The effectiveness of Voluntary Sector Organisations in comparison with the National Health Service: Patient profiles, therapist effects & treatment outcomes

Keeley A. Rolling

Submitted for the award of Doctor of Clinical Psychology

University of Sheffield

June 2016

Access to Thesis

The University Of Sheffield.

Access To Thesis.

A fully completed copy of this form must be submitted to Research & Innovation Services prior to the award of your degree. If you are submitting a hard copy of the thesis the form should be bound into the front of the thesis

SECTION 1: STUDENT DETAILS			
Family Name	Rolling	First Name	Keeley
Registration Number	120106165	Department	Clinical Psychology
Thesis Title	The effectiveness of Voluntary Sector Services in comparison with the NHS: Patient profiles, th		

SECTION 2: THESIS SUBMISSION DETAILS – PLEASE SELECT ONE OF THE FOLLOWING OPTIONS	
☐	I am submitting in print format only for deposit in the University Library **(Note: this option only applies to students who initially registered prior to 2008)**
☐	I am submitting an eThesis only to the White Rose eTheses Online server. I confirm that the eThesis is a complete version of my thesis and no content has been removed
☑	I am submitting an eThesis to the White Rose eTheses Online server and also submitting in print format because I have removed some content from my eThesis

SECTION 3: EMBARGO DETAILS – PLEASE SELECT FROM THE FOLLOWING OPTIONS

Each Faculty has agreed a pre-approved embargo threshold (Arts & Humanities – 1 yr; Engineering – 1 yr; Medicine, Dentistry & Health – 2 yrs; Science – 5 yrs; Social Sciences – 3 yrs. Requests for embargoes that exceed the Faculty threshold will require Faculty approval. If you wish to request a longer embargo, please complete and submit the form available at: www.shef.ac.uk/ris/pgr/code/embargoes

Please note that if no boxes are ticked, you will have consented to your thesis being made available without any restrictions.

Should the thesis be embargoed? If 'Yes', please specify the length of embargo requested (in years)	Print Thesis	☐ No	☑ Yes	5 Years
	eThesis	☐ No	☑ Yes	5 Years

Reason for the embargo (please select from the following options):

☐	Third party copyright	☐	Commercial confidentiality
☐	Contains personal data	☐	Could prejudice national security
☐	Could endanger health and safety	☐	Exempt under another category listed in the FOI Act 2000
☑	Planned publication	☐	Other

SECTION 4: COPYRIGHT LICENCE OPTIONS – PLEASE SELECT ONE OF THE FOLLOWING

This thesis is protected by the Copyright Design and Patents Act 1988. No reproduction is permitted without consent of the author. It is recommended that you make your thesis available using a Creative Commons Licence http://creativecommons.org/about/license/. This Licence protects you as the author of the work and also clarifies the uses that others may make of your work.

☑	Creative Commons Attribution-Non-Commercial-No-derivatives **(recommended)**	☐	Creative Commons Attribution-Non-Commercial
☐	Creative Commons Attribution	☐	Creative Commons Attribution-No-derivative-Works
☐	Creative Commons Attribution-Non-Commercial-Share Alike	☐	Other/Do not apply a Licence

SECTION 5: THESIS DEPOSIT AGREEMENT - STUDENT		
1. I, the author, confirm that the Thesis is my own work, and that where materials owned by a third party have been used, copyright clearance has been obtained. I am aware of the University's *Guidance on the Use of Unfair Means* (www.sheffield.ac.uk/ssid/exams/plagiarism).		
2. I confirm that all copies of the Thesis submitted to the University, whether in print or electronic format, are identical in content and correspond with the version of the Thesis upon which the examiners based their recommendation for the award of the degree (unless edited as indicated above).		
3. I agree to the named Thesis being made available in accordance with the conditions specified above.		
4. I give permission to the University of Sheffield to reproduce the print Thesis (where applicable) in digital format, in whole or part, in order to supply single copies for the purpose of research or private study for a non-commercial purpose. I agree that a copy of the eThesis may be supplied to the British Library for inclusion on EThOS and WREO, if the thesis is not subject to an embargo, or if the embargo has been lifted or expired.		
5. I agree that the University of Sheffield's eThesis repository (currently WREO) will make my eThesis (where applicable) available over the internet via an entirely non-exclusive agreement and that, without changing content, WREO and/or the British Library may convert my eThesis to any medium or format for the purpose of future preservation and accessibility.		
6. I agree that the metadata relating to the eThesis (where applicable) will normally appear on both the University's eThesis server (WREO) and the British Library's EThOS service, even if the eThesis is subject to an embargo.		
Student's name (PLEASE PRINT): K Rolling	Signature:	Date 23/05/2016
SECTION 6: THESIS DEPOSIT AGREEMENT - SUPERVISOR		
I, the supervisor, agree to the named Thesis being made available in accordance with the conditions specified above.		
Supervisor's name (PLEASE PRINT): andrew thompson	Signature: 	Date: 23/05/2016

SECTION 6: TO BE COMPLETED BY RESEARCH & INNOVATION SERVICES		
Does the embargo exceed the agreed Faculty length? FCA - 1YR; FCE - 1YR; FCM - 2YRS; FCP - 5 YRS; FCS - 3 YRS	☐ Yes* if 'yes' please attach embargo extension request form	☐ No
University stamp		

Declaration

I confirm that this thesis has not been submitted for any other degree or to any other institution

Structure

Literature Review: 8, 131 (excluding references)

11, 049 (including references & appendices)

Empirical Report: 11, 997 (excluding references)

15, 398 (including references & appendices)

Thesis Total: 20, 128 (excluding references)

26, 447 (including references & appendices)

Thesis Abstract

Literature Review: Twenty-two practice-based evidence studies using the CORE-OM were reviewed and patient outcomes summarised according to mental health provider. Studies using the CORE-OM could be categorised according to four main areas of evaluation, including global service effects, intervention effects, and contextual factor effects. Indications of factors influencing outcomes were considered. Treatment outcomes across mental health services averaged an uncontrolled effect size of 1.23 and RCSI rate of 41.5%. Evidence suggested a range of mental health providers were effective in reducing psychological distress and highlighted understanding both therapeutic change and service effectiveness requires a multifaceted approach. Further research is indicated to develop evidence concerning services beyond primary care and the synergistic effects of multiple factors on outcomes.

Empirical report: Secondary analyses of a large practice-based dataset was undertaken from which to determine and compare service profiles, predictors of therapeutic outcome, therapist effects among voluntary sector organisations (VSOs) and national health service (NHS) mental health providers. The sample comprised a total of 10,142 patients and 133 therapists. VSOs and NHS providers were primarily differentiated by their therapeutic orientations, treatment durations, and therapist caseload sizes. Outcomes between sectors were broadly comparable, with recovery rates of 58% and 61% for VSO and NHS providers respectively. VSOs had therapist effects of 4.5%, considerably smaller than therapist effects among the NHS sector (12.7%). Common and specific predictors of outcomes between sectors were identified and explored, with implications for clinical practice and future research discussed.

Acknowledgements

I would like to thank my thesis supervisor, Professor Michael Barkham, and collaborator, Dave Saxon, for their support, guidance, and expertise throughout the project. A special thank you to Dr. Andrew Thompson for all your help during the final stages and toward submission.

To my Mum and Dad, Nicola and Alicia, I owe you my deepest gratitude for your unwavering support, belief, patience, and encouragement throughout the training process. Thank you for never giving up on me and always being there.

To Nannan Rolling and Grandad Hartley, thank you for giving me strength and integrity - may you forever be at peace.

Thank you to all of my friends for never quite getting it but being there anyway.

Contents

Empirical Report: The effectiveness of Voluntary Sector Organisations in comparison with the National Health Service: Patient profiles, therapist effects & treatment outcomes

Appendices

<Page left blank>

Literature Review

A Systematic Review of Practice-Based Evidence and Service Outcomes derived from the Clinical Outcomes in Routine Evaluation – Outcome Measure

Abstract

Objectives: Uniquely focused on the CORE-OM, the review aimed to critically evaluate Practice Based Evidence (PBE) studies and summarise patient outcomes according to mental health provider.

Methods: Databases, *Scopus*, *Web of Science*, and *PsychInfo*, were systematically searched using the terms "Clinical Outcomes in Routine Evaluation-Outcome Measure", "Clinical Outcomes in Routine Evaluation Outcome Measure", "CORE-OM", and "CORE OM". Studies meeting specified inclusion criteria were subject to quality appraisal. Primary outcomes extracted concerned effect sizes and rates of reliable and clinically significant improvement (RCSI).

Results: Twenty-two studies drawn from National Health Service and alternative mental healthcare providers comprised the review. Studies using the CORE-OM could be categorised according to four main areas of evaluation, including global service effects, intervention effects, and impact of contextual factors. Treatment outcomes across mental health services averaged an uncontrolled effect size of 1.23 and RCSI rate of 41.5%.

Conclusion: A range of mental health services are effective in reducing psychological distress. However, multiple factors contribute toward successful outcomes and there is a need for greater synergy of such factors in order that effectiveness be fully and accurately determined.

Practitioner Points

Clinical Implications:

- A range of counselling and psychological services are effective in improving patient distress.
- Multiple factors are potentially associated with treatment outcomes and require further investigation in order that understandings of psychotherapeutic change are improved.

Limitations:

- Study heterogeneity prevented formal quantitative pooling.
- Methodological weaknesses of studies, such as inadequate consideration of confounds and under-reporting, potentially inflated treatment outcomes.

Introduction

Traditionally, evidencing psychological treatments has relied upon *efficacy* studies using Randomised Controlled Trial (RCT) methodology. Such studies report outcomes of treatments delivered to highly select groups of participants under controlled and manualised conditions. Due to these characteristics, RCTs are associated with high internal validity and have often been considered 'gold standard' means of addressing questions of therapeutic effects. Treatments subject to RCT methodology with demonstrable effects are commonly described as empirically validated evidence-based therapies and are organised within clinical practice guidelines. Meta-analytic data reports moderate to large effect sizes across therapeutic models and mental health conditions, with broad equivalence of interventions emerging (e.g., Smith & Glass, 1977). Such findings have offered support to global conclusions regarding the beneficial effects of psychological therapies.

However, questions remain regarding the extent to which observed treatment effects derived from *efficacy* studies are obtainable in routine practice (Seligman, 1995). In order to address these concerns, Practice Based Evidence (PBE) research has been advocated as a complimentary approach (Barkham, Hardy, & Mellor-Clark, 2010). PBE evaluates treatments in 'real-world' clinical settings from which to determine their *effectiveness* whilst accounting for contextual factors, such as patient symptom severity, that might impact treatment effects. Thus, PBE research offers greater external validity than *efficacy* studies and is necessary if the impact of psychotherapy is to be fully understood. Furthermore, the methods employed within PBE, such as monitoring patient outcomes and establishing performance indicators, provide the means for services to examine the quality of their care and be responsive to areas requiring

improvement in keeping with NHS quality frameworks (e.g., Department of Health, [DoH], 2004).

Despite mandates for increased evidencing of clinical practice (DoH, 2004) few studies pertaining to the systematic review and synthesis of PBE research exist. Initial attempts to address questions of 'real-world' outcomes of psychological therapies have concerned the secondary analyses of meta-analytic data (Shadish et al., 1997; Shadish, Matt, Navarro, & Phillips, 2000). These analyses rated 56 and 90 studies, respectively, according to continuums of clinical representativeness. Findings suggested that while greater degrees of clinical representativeness were, descriptively, associated with smaller magnitudes of therapeutic effect, the differences were not statistically significant. The authors concluded psychological therapies having established effectiveness in both research and clinical settings. However, the degree to which studies determined 'clinically representative' constituted PBE is unclear.

A more recent meta-analytic study, focusing exclusively on 31 PBE studies published between 1990 and 2008, compared effect sizes and Reliable and Clinically Significant Improvement (RCSI) rates with efficacy benchmarks (Cahill, Barkham, & Stiles, 2010). Fixed effects analysis of 14 included studies yielded effect sizes of 1.29, .79 to 1.08, and 1.60 for presentations of common mental health problems (CMHPs), panic disorder, and bulimia nervosa, respectively. A RCSI rate of 56% was available for CMHPs only. The authors concluded that, while effect sizes were lower than derived efficacy benchmarks, patients experiencing a range of mental health disorders benefitted from psychological interventions delivered in routine practice with equivalent rates of RCSI (54%) achieved. However, a number of limitations associated with the Cahill et al. (2010) review were identified: i.) included studies were limited in their scope, generally being restricted to depression and primary care; ii.) included studies

employed a range of outcome tools, introducing issues of measurement reactivity (Smith et al., 1980) and differing RCSI thresholds, rendering RCSI estimates questionable; and iii.) contextual effects were not accounted for, receiving limited attention, which should be a focus of PBE research.

Rationale and Objectives

The limited number of systematic reviews of PBE research suggests the state of evidence of psychotherapeutic outcomes continues to be dominated by *efficacy* based studies. Moreover, reviewing evidence of psychological therapies delivered in routine practice is emphasised due to quality assurance frameworks placing greater impetus on mental health providers to evaluate their practices and systematically assess patient outcomes (e.g., DoH, 2012; 2014). Together, these factors suggest there remains scope to improve current understanding of psychotherapy in routine practice and a timely review of PBE outcomes is warranted.

In recognition of the issues associated with the Cahill et al. (2010) review outlined above, the current review focused on a single outcome measure: the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM: Barkham et al., 1998, 2005). Briefly, the CORE-OM comprises 34 self-report items of psychological distress and encompasses domains of subjective well-being, symptoms (anxiety, depression, physical problems, and trauma), functioning (general functioning, close relationships, and social relationships) and risk (risk to self and risk to others). Each item is scored on a 5-point scale, anchored 0 '*not at all'* to 4 '*most or all of the time*'. Clinical scores range between 0 and 40, with higher scores indicative of greater levels of distress. Recommended cut-off scores have been established: <10 low level (non-clinical), 10-14 mild, 15-19 moderate, 20-24 moderate-severe, and ≥25 severe. The CORE-OM has established psychometric properties, with an internal consistency of .94 (Barkham et al.,

2001) and test-retest correlations of ≥.80 (Barkham, Mullin, Leach, Stiles, & Lucock, 2007).

The CORE-OM was developed to be applicable across psychological disorders and settings, is free to use, and designed to facilitate sustainable and meaningful use of outcomes in routine practice. Since the development of the CORE-OM, approximately 20 years ago, it has become one of the most widely used tools among mental health providers (Jacobs, 2009). However, a comprehensive review of the evidence yielded from the implementation of the CORE-OM has not been completed. It is for these purposes that the CORE-OM represented a suitable candidate for review.

The specific aims of the review were to: i.) systematically identify and critically evaluate PBE studies using the CORE-OM as the primary outcome tool; and ii.) summarise patient outcomes, measured by the CORE-OM, according to mental health service provider.

Method

Search Strategy

A systematic search was undertaken, between June and August 2015, of three databases: Scopus, Web of Science, and PsychInfo. Search terms included "Clinical Outcomes in Routine Evaluation-Outcome Measure", "Clinical Outcomes in Routine Evaluation Outcome Measure", "CORE-OM", and "CORE OM", targeting titles, topics, and keywords. Reference lists of eligible studies were searched for additional records potentially missed by electronic searches.

Selection Procedure

Returned results yielded 385 publications. Following removal of duplicates and screened titles and abstracts that were not appropriate, 137 records were assessed according to inclusion/exclusion criteria. Hand searches of reference lists provided a further two records. A total of 22 studies were yielded for review. Study selection processes are depicted diagrammatically in Figure 1 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA] diagram; Moher, Liberati, Tetzlaff, & Altman, 2009).

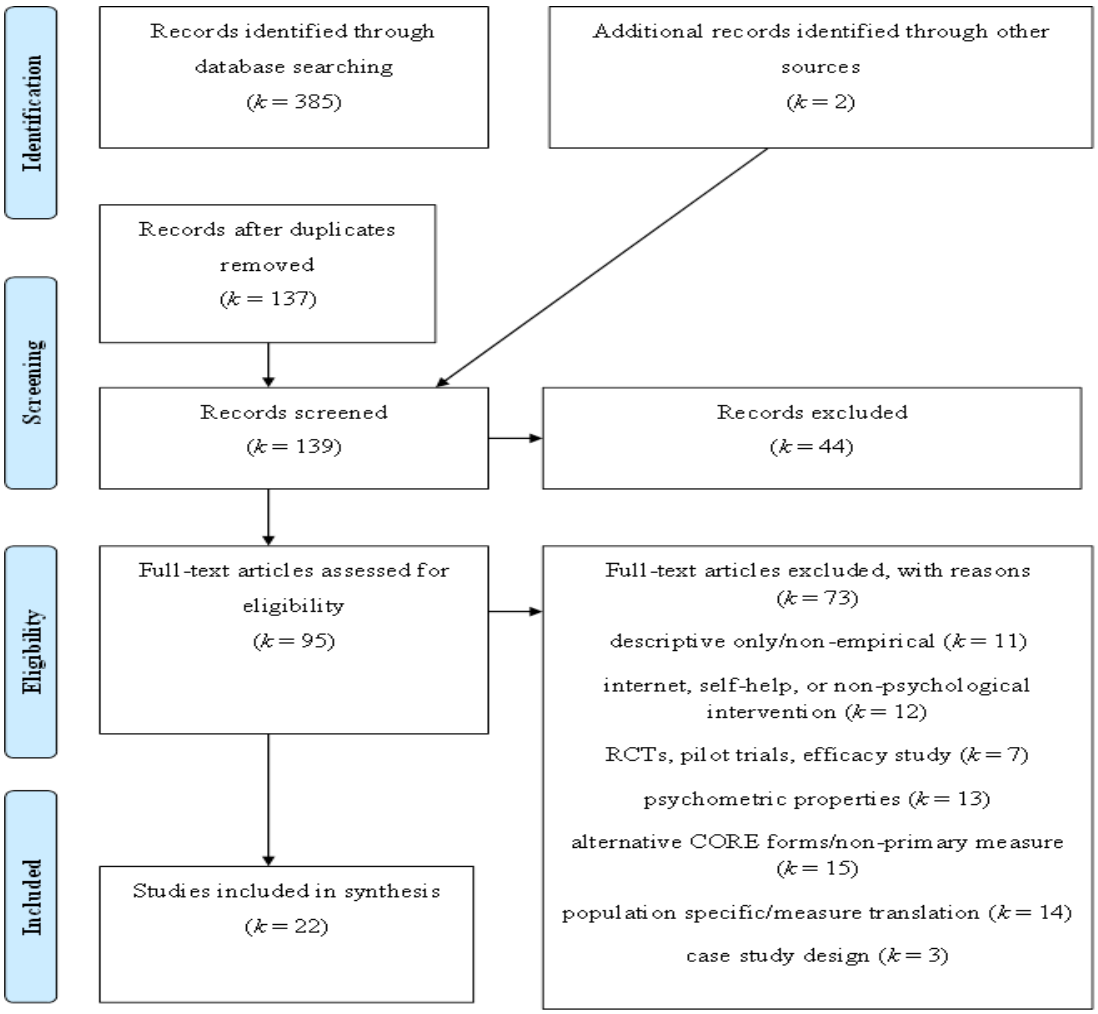

Figure 1. PRISMA diagram of study inclusion strategy

Study Selection Criteria

Inclusion criteria for studies were as follows: i.) publication in a peer-reviewed journal and in English language; ii.) report empirical data; iii.) use an adult sample; iv.) evaluation of the effectiveness of a component of service delivery using the CORE-OM as the primary measure[1]; and v.) be considered practice-based evidence research. For the purposes of the current review, a broad definition of service delivery was adopted in order to reflect the multifaceted nature of implementing mental health care. Levels of service delivery included, but were not limited to, tiers or types of providers, psychological interventions and intervention delivery, and therapist factors. Studies of measure development were permitted where outcome data were reported and able to be extracted.

Primary exclusion criteria were: i.) studies evaluating computer based interventions; ii.) Randomised Controlled Trials (RCTs); and iii.) case study designs. The purpose of these exclusion criteria was to reflect the scope of CORE-OM use in routine practice for which it was designed and to restrict inclusion of lower order quality studies based on a single subject. Study quality was not used further to determine inclusion or exclusion into review.

Treatment of Duplicate Data

Duplicate data were not permitted unless the data were used for different purposes in keeping with the review aims and inclusion criteria. Where studies used the same dataset for the same purposes, the study included for review represented the most recent use of the data. As such, one study (Stiles, Barkham, Twigg, Mellor-Clark, &

[1] Studies using a battery of measures of general psychological well-being, and therefore a primary measure was not defined, were included and data derived from the CORE-OM extracted.

Cooper, 2006) was removed from the review due to a more recent use of the same dataset.

Quality Assessment

Assessment of study methodological quality was determined using the Downs and Black (1998) checklist. The checklist is suitable for both randomised and non-randomised studies. For the purpose of the current review, and to be in keeping with previous practice-based evidence reviews, an adapted version of the checklist was used (Appendix A: Cahill et al., 2010) and item 28 was rated 0 or 1 (rather than 0 to 5). Studies were rated on 28 items, yielding a total score and scores on four dimensions: study reporting, external validity, internal reliability, and bias. Study quality was described in accordance with conventions described by Samoocha et al. (2010); ≤14 'poor', 15-19 'fair', 20-25 'good', and ≥26 'excellent'.

A random sample of five studies was second rated by an independent other (Trainee Clinical Psychologist). Agreement level among ratings was k = .78, placing agreement within the 'good' category according to Landis and Koch (1977) classification. Inconsistencies in ratings were resolved through discussion and re-examination of individual ratings.

Data Extraction and Synthesis

Characteristics of study designs, populations, treatment and intervention processes, and service delivery components were examined. Formal meta-analyses were not undertaken due to the heterogeneity of studies and their samples. Rather, studies were reported individually according to the factor(s) under investigation, with reference to their specific design features, methodological quality, and results. For the synthesis of patient outcomes per service, effect sizes and rates of Reliable and Clinically Significant

Change (RCSC) were extracted directly from studies or calculated from available study data where possible. Four RCSC criteria were considered:

i.) *Reliable and Clinically Significant Improvement* (Recovered): comprising patients whose pre-post scores met *study* criteria for both reliable improvement and clinically significant improvement.

ii.) *Reliable Improvement*: consisting patients whose pre-post score difference met *study* criteria for reliable improvement only.

iii.) *No Reliable Change*: referring to those patients whose pre-post CORE-OM scores showed no movement or movement that did not reach *study* reliable change threshold.

iv.) *Deteriorated*: comprising patients who had post therapy scores reliably worse than their intake scores.

Results

Descriptive Overview of Studies

Study characteristics, alongside quality ratings, are summarised in Table 1. The majority of studies were conducted in the United Kingdom, with one study taking place in Sweden. All studies had been published since 2000. The majority of studies (k = 12) reported data from primary care services. Two studies were from secondary care services, 3 studies were drawn from specialist services, 4 studies were from alternative mental health providers, and 1 study reported data from multiple settings.

In consideration of the service delivery components under investigation, there was a range in study focus: 8 studies focused on the effectiveness of a specific

intervention and 2 studies compared the effectiveness of different types of treatments, 6 studies were characterised by providing data concerning the effectiveness of a particular care setting or provider, 3 studies concerned the effects of process factors such treatment dose or waiting times, and 3 studies investigated therapist effectiveness and therapeutic conditions. These factors were examined using three primary methodological designs comprising pre-post single group design (k =11), aggregated design (k = 9), and non-equivalent groups design (k = 2).

Sample Characteristics per Provider

Given that mental health services are designed to meet the needs of, or are accessed by, potentially diverse groups of patients, study sample characteristics were described in relation to the type of service provider. The purpose was to compile profiles of patients presenting to different types of mental health services, based on study information, in order to inform understandings of treatment and service outcomes.

A degree of patient profile data was available across studies. However, comprehensiveness and quality of patient data provided was variable, with data concerning ethnicity, employment status, and relationship status being the least routinely reported information. Additionally, there were inconsistencies in the use of samples from which to derive profile data, with some studies providing data for both *referred* and *completer* samples and others reporting data of only one sample type. In order that summary data were not skewed by reporting differences of individual studies, summary statistics of sample sizes were described using the median. One study (Stiles, Barkham, & Wheeler, 2015) was excluded from the profile data synthesis due to the sample profile comprising aggregated patient data from multiple care settings.

Primary Care. Samples ranged between 112 and 11,953 patients (*Mdn* = 1,868, k = 12). Females were found to account for the largest proportion of all samples,

ranging between 56.5% and 76.0%, and the majority of patients were White British/European (range = 84.5 - 95.0%, *k* = 7). Patient average age was 40.3 years (SD = 1.8; range = 37.3 – 42.8 years; k = 8).

Illness characteristics were available from eight studies, with one study (Houghton & Saxon, 2007) excluded due to a diagnosis of anxiety being the primary inclusion criteria. The eight studies identified patients attending primary care commonly presenting with depression (range = 34 - 78.0%, *k* = 8) and anxiety (range = 28 - 84.6%, *k* = 8). Average pre-therapy CORE-OM scores fell within the moderate severity level, ranging between 16.78 and 19.30 (*Mean* = 18.1, *SD* =.8, *k* = 10).

Secondary Care. A less comprehensive profile of secondary care patients was able to be drawn due to the limited number of available studies and under-reporting within included studies. Patient numbers were 224 (Beck, Burdett, & Lewis, 2015) and 2,710 (Barkham et al., 2001) for *referred* samples. Similar to primary care studies the largest proportions of patients were female, 60.2% and 63.0% respectively. Beck et al. (2015) reported a mean age of 41.0 years for the *referred* patient sample, which resembled the *completer* subsample of the Barkham et al. (2001) study (*M* = 40.9 years, SD = 15.2, *n* = 224).

Beck et al. (2015) reported their sample as predominantly White (68.2%) and common presentations of mood/affective disorders (35.2%) and neurotic disorders (39.3%). These data were not available from the Barkham et al. (2001) study. Differences were observed in pre-treatment score, which averaged 21.5, between *completer* subsamples. Barkham et al. (2001) reported average scores of 18.5 and Beck et al. (2015) reported median scores of 24.4, reflecting differing levels of severity at intake. However, the higher levels of distress observed by Beck et al. (2015) are representative of patients from a single secondary care service purposefully sampled for

study. In contrast, Barkham et al. (2001) aggregated data from six secondary care sites and reported levels of distress are therefore more likely to represent the variability in the severity of patient conditions who present to secondary care services.

Specialist Services. Three studies were considered from specialist care providers. These studies evaluated outcomes of interventions for specific subsets of patients, including those diagnosed with Borderline Personality Disorder (BPD; McFetridge & Coakes, 2010), functional neurological disorders (Reuber, Burness, Howlett, Brazier, & Grunewald, 2007), and those described as experiencing enduring and chronic mental health difficulties (Paley et al., 2008). Sample sizes were relatively small, particularly in comparison with primary and secondary care studies, ranging between 40 and 67 patients. However, these sample sizes are likely to reflect the nature of specialist services in terms of demand and use by smaller subsets of patients. Two studies (Paley et al., 2008; Reuber et al., 2007) included males and females in their samples, with females accounting for the highest proportion of patients (74.6% and 81.0%, respectively). McFetridge and Coakes (2010) recruited only female patients.

Patient age ranged between 31.5 and 44.2 years (*Mean* = 37.2 years, *Mdn*=36.0 years, *k*=3). Only one study (Reuber et al., 2007) provided further demographic information of their sample, which was limited to employment status, comprising 67.7% unemployed or patients receiving benefit payments. Reporting of pre-treatment scores was variable and inconsistent across studies, with Reuber et al. (2007) reporting mean total scores of 51.1, Paley et al. (2008) observed values of 20.3, and McFetridge and Coakes (2010) reported a median value of ≥25. Such values place specialist care patient scores within moderate, moderate to severe, and severe ranges. Differences in average intake severity scores may be attributable to the types of patients recruited, particularly in terms of their presenting problems and associated levels of risk.

Alternative Providers. Of the four studies from alternative mental health providers, two reported data from voluntary sector services (Armstrong, 2010; Hodge, Barr, Bowen, Leeven, & Knox, 2012) and two reported data from university counselling services (Connell, Barkham, & Mellor-Clark, 2008; Murphy & Cramer, 2014). Sample sizes varied between 35 and 323 patients across studies (*Mdn* = 95) and consisted primarily of female patients; ranging between 65.7% and 77.0%.

Armstrong (2010) reported a mean age of 36.0 years and found 55.0% of patients were employed, 40.0% were unemployed, and 58.0% had received no prior therapeutic input. Hodge et al (2012) described patients as retired, with a mean age of 58.9 years. University samples averaged 22.8 to 24.3 years. Samples were described as predominantly White/European (range = 79 – 97.0%). Depression, anxiety, and interpersonal relationship difficulties were found as the most common presenting problems. Similar to primary care studies, patients presenting to alternative providers were found to have pre-therapy scores within the moderate severity range (*Mdn*=18.1, *k*=4).

Table 1. *Study Characteristics*

Lead Author	Year	Design	Evaluative Focus	N	Sample Characteristics	Interventions	Quality Rating
Primary Care							
Andrews	2011	Pre-Post	Treatment Effect	124	71.0% female, mean age = 42.8 years (SD = 12.7); patients presented with anxiety and depression; moderate intake severity	Human Givens	24
Barkham	2006	Aggregated	Dose Effect	1868	73.1% female, mean age = 40 years (SD = 12.8); primary presenting problems were anxiety and depression; moderate intake severity	Common approaches were PCT, Integrative, CBT	20
Barkham	2012	Aggregated	Care Setting	9761	72.3% female, mean age = 40.8 years (SD = 12.8); primary presenting problems were anxiety, depression, and interpersonal difficulties; moderate intake severity	Common approaches comprised Integrative, PCT, Structured/Brief	21
Evans	2002	Aggregated	Care Setting	125	Moderate intake severity	-	15
Evans	2003	Aggregated	Care Setting	6610	71.5% female, mean age = 38 years (SD = 13.1); presenting problems of anxiety, depression, and interpersonal difficulties; moderate intake severity	-	19

Lead Author	Year	Design	Evaluative Focus	N	Sample Characteristics	Interventions	Quality Rating
Gibbard	2008	Pre-Post	Treatment Effect	1098	72% female, mean age = 40.5 years; common presenting problems were anxiety and depression; moderate intake severity	PCT	19
Holmqvist	2014	Non-equivalent groups	Treatment Types	733	Overall sample were 76.0% female, mean age = 37.3 years (SD = 14.3); moderate intake severity	Three groups: reflective, directive, and supportive	20
Houghton	2007	Pre-Post	Treatment Effect	191	56.5% female; patients experiencing anxiety recruited	Psycho-educational CBT Group	14
McHugh	2013	Pre-Post	Treatment Effect	43	Patients with mild to moderate disorders recruited, commonly depression and anxiety	Individual brief CBT	12
Mullin	2006	Aggregated	Care Setting	11593	72.1% female; presenting problems were anxiety, depression, and interpersonal difficulties; moderate intake severity	-	20

Lead Author	Year	Design	Evaluative Focus	**N**	Sample Characteristics	Interventions	Quality Rating
Saxon	2012	Aggregated	Therapist Effects	10786	71.5% female; presented with problems of anxiety and depression; moderate intake severity	-	22
Stiles	2008	Non-equivalent groups	Treatment Types	5613	70.7% female, mean age = 40.7 years (SD = 12.7); presenting problems were anxiety, depression, and interpersonal difficulties; moderate intake severity	Six groups: CBT, PCT, PDT, CBT+1, PCT+1, PDT+1	23
Secondary Care							
Barkham	2001	Aggregated	Care Setting	224	61.6% female, mean age = 40.9 years (SD = 15.2); moderate intake severity	-	20
Beck	2015	Pre-Post	Waiting Times	224	60.2% female, mean age = 41 years; presenting problems were mood and neurotic disorders, moderate to severe	-	20
Specialist Care							
McFetridge	2010	Pre-Post	Treatment Effect	40	Female only sample diagnosed with borderline personality disorder, mean age = 31.5 years, severe intake severity	DBT informed	16
Paley	2008	Pre-Post	Treatment Effect	67	74.6% female, mean age = 36 years (SD = 10); presenting problems described as severe and enduring; moderate to severe intake severity	PIT	19

Lead Author	Year	Design	Evaluative Focus	**N**	Sample Characteristics	Interventions	Quality Rating
Reuber	2007	Pre-Post	Treatment Effect	94	81.0% female, mean age = 44.2 years (SD = 12.6); patients experiencing functional neurological symptoms recruited; moderate intake severity	PIT	17
Alternative Provider							
Armstrong	2010	Pre-Post	Therapist Effects	118	71.0% female, mean age = 36 years; common presenting problems were depression, interpersonal difficulties, and anxiety; moderate intake severity	Solution focused	18
Connell	2008	Aggregated	Care Setting	323	70.5% female, mean age = 23.1 years (SD=6.3); common presentations of anxiety, interpersonal difficulties, and depression; moderate intake severity	Primarily psychodynamic, brief/structured, integrative;	22
Hodge	2012	Pre-Post	Treatment Effect	35	65.7% female, mean age = 58.9 years; recruitment of patients diagnosed with visual impairments; moderate intake severity	Emotional Support and Counselling	14
Murphy	2014	Pre-Post	Therapeutic Conditions	72	77.0% female, mean age = 24.3 years (SD = 8.6); moderate intake severity	Humanistic and CBT	19

Lead Author	Year	Design	Evaluative Focus	N	Sample Characteristics	Interventions	Quality Rating
Stiles	2015	Aggregated	Dose Effect	26430	69.3% female, mean age = 38.6 years; presenting problems were commonly anxiety, interpersonal difficulties, and depression; moderate intake severity	Common approaches were integrative, PCT, and Psychodynamic	

Note. CBT = Cognitive Behavioural Therapy; PCT = Person-Centred Therapy; PDT = Psychodynamic Therapy; DBT = Dialetic Behaviour Therapy; PIT = Psychodynamic Interpersonal Therapy

Study Quality

Of the 22 studies, 10 studies were of good quality, 9 were of fair quality, and 3 were poor quality. No included studies were rated as 'excellent' quality. The average quality rating of studies fell within the fair range (M = 18.73, SD = 2.98) and ranged between 12 and 24.

In terms of quality according to study design, the proportions of quality items met by studies are presented in Table 2 (individual item ratings are available in Table 3; Appendix B). Data showed that studies had high levels of reporting but were often characterised by lower levels of quality concerning internal validity bias.

Table 2.

Percentage of quality items achieved by included studies according to design

Study Design (n)	Dimension				Overall (%)
	Reporting (%)	External Validity (%)	Internal Reliability (%)	Bias (%)	
Aggregated (9)	74.7	65.7	64.4	31.1	63.2
Non-Equivalent (2)	89.5	68.2	70.0	40.0	67.2
Single Groups (11)	73.6	62.0	63.6	27.3	53.1

Pre-post single group designs (k = 11) constituted the majority of evidence and were found to achieve lower proportions of items across the four quality dimensions and overall than other designs. Varying degrees of methodological quality among pre-post design studies was observed (M = 17.45, SD = 3.36, range = 12 – 24), with 7 studies of fair quality and a single study being of good quality. All poor quality studies (k = 3) used pre-post designs.

Methodological quality among aggregated designs (k = 9) was less variable than pre-post designs (M = 20.22, SD = 2.33, range = 15 – 23) and generally fell within the good range (k = 7). The remaining two studies were considered fair quality and were the earliest publications of such designs extracted. Non-equivalent groups design was adopted by two studies, both of which were of good quality (M = 22, SD = 1.41, range = 21 – 23) and demonstrated higher proportions of items achieved overall and across dimensions.

Narrative Synthesis

Service Provider Effects. Six studies investigated global therapeutic change at the service level using aggregated designs. Studies were characterised by the use of large practice-based datasets. Four studies were drawn from primary care (Barkham, Stiles, Connell, & Mellor-Clark, 2012; Evans et al., 2002; Evans, Connell, Barkham, Marshall, & Mellor-Clark, 2003; Mullin, Barkham, Mothersole, Bewick, & Kinder, 2006), one from secondary care (Barkham et al., 2001), and one from university counselling services (Connell et al., 2008). The quality of these studies was generally good, with only two studies (Evans et al., 2002; Evans et al., 2003) rated as fair quality. These latter studies were weaker in their general reporting, particularly concerning the characteristics of their samples and patient selection processes.

Within primary care settings, a degree of variation in the outcomes observed. Evans et al. (2002) investigated the outcomes of a relatively small sample of 124 patients; reporting significant improvements in psychological distress, with 49% of patients achieving RCSI. While the outcomes were suggestive of positive effects, Evans et al. (2002) failed to provide information concerning the types of statistical tests used or the values produced from their analyses. In a larger study of improved quality comprising 6,610 patients from 33 sites, Evans et al. (2003) found 77.8% of patients

achieving reliable improvement or clinically significant change. Given that change categories were not differentiated in the latter study, the outcomes are somewhat misleading and lead to the potential for overestimation of service effects.

Mullin et al. (2006) established primary care service benchmarks for the proportions of patients meeting reliable and clinically significant change criteria. The sample was 11,953 patients of 32 sites who presented with multiple mental health difficulties, primarily depression and anxiety. Benchmarking analyses indicated an average RCSI rate of 55% across services. Additional analyses of therapist level (n = 513) benchmarks, adjusting for patient case-mix severity, revealed average RCSI rates of 63% and 38% for 'non-severely' distressed and severely distressed patients respectively. These findings suggested that RCSI rates are impacted by patient intake severity, with those patients experiencing greater levels of distress being less likely to achieve RCSI status post-therapy. Similar RCSI rates were observed by Barkham et al. (2012) who reported 58.3% of patients, from a sample of 9,761 patients and 34 sites, meeting criteria for RCSI and a large pre-post effect size of 1.79.

For primary care studies assessing service outcomes using practiced based datasets, the proportions of patients achieving RCSI were generally in the region between 50 and 60%, with higher quality studies (Barkham et al., 2012; Mullin et al., 2006) showing greater consistency in findings. Potentially, the differences in outcomes are reflections of natural variation across these types of services. However, factors associated with study methodological quality, particularly in terms of the representativeness of data, might also account for differing findings.

In terms of secondary care, Barkham et al. (2001) established service effects from a dataset of 224 patients from six sites. Patients represented a subsample of completers, who had pre and post-treatment data available, drawn from a larger pool of

2,710 patients within 39 secondary care services. The authors reported a large pre-post treatment effect size of .87 and 39% of patients meeting RCSI criteria. Greater rates of RCSI were also found among patients with moderate levels of psychological distress than patients whose scores fell within the severe range. Thus, similar to the primary care study by Mullin et al. (2006), intake severity was associated with poorer outcomes. Despite checks for representativeness, which revealed no significant differences between completer and full intake samples regarding severity, the completer sample comprised just 8% of patients rendering interpretation and generalisation of the results questionable.

Connell et al. (2008) aggregated data from 11 university counselling services comprising 1,189 patients. A range of therapies were used, with the most commonly received intervention comprising psychodynamic, brief/structured, and integrative approaches delivered across an average of four sessions. Following application of exclusion criteria concerning incomplete and unreliable data, the final sample subject to analyses comprised 323 patients from 7 services. Outcomes reported were 54.3% of patients scoring above clinical cut-off at intake achieving RCSI and a large pre-post effect size of 1.57. While the overall study quality was good, caution is warranted due to potential selection bias and the low rate of complete data (reported as 38.2%).

The findings drawn from the above studies suggest variation in outcomes between types of mental health services. However, while the use of practice datasets provide a means of analysing large samples, data are likely to be collected under different conditions that are not measured and/or reported, thus representing unknown variations in outcomes. For instance, given the focus on global service effects, it is not possible to delineate distinguishing features of services, interventions, or processes that contribute toward outcomes. Moreover, while the impact of case-mix factors on

outcomes was not a common feature among studies, with only two studies examining intake severity profiles, the consistency of the finding that more severely distressed patients are less likely to achieve RCSI suggests patient case-mix factors are potential sources of variation in outcomes. Without further consideration of patient level factors, established outcomes could be over-estimated and lead to inappropriate conclusions regarding one service type being more effective than another. Additionally, there were frequent issues across studies concerning low completion rates, which were not always clearly reported or addressed, and associated implications for generalisation.

Intervention Effects. The effectiveness of interventions was the focus of ten studies. Interventions investigated were diverse in respect of the types investigated, either evidence-based or under-represented treatments, their delivery formats, and recruited samples.

The majority of intervention studies (k = 8) used pre-post single group designs, which were characterised by evaluating patient outcomes in one service, a specific subset of patients, or of a discrete intervention. A number of studies involved a combination of the above. For these eight studies, data were reported from primary care (k = 4; Andrews, Twigg, Minami, & Johnson, 2011; Gibbard & Hanley, 2008; Houghton & Saxon, 2007; McHugh, Brennan, Galligan, McGonagle, & Byrne, 2013), specialist services (k = 3; McFetridge & Coakes, 2010; Paley et al., 2008; Reuber et al., 2007), and a voluntary provider (k = 1; Hodge et al.,2012). The remaining two studies were non-equivalent group designs determining the comparable effectiveness of multiple interventions (Holmqvist, Ström, & Foldemo, 2014; Stiles, Barkham, Mellor-Clark, & Connell, 2008).

Primary care studies described CBT approaches (k = 2: Houghton & Saxon, 2007; McHugh et al., 2013), person-centred therapy (PCT; k = 1: Gibbard & Hanley,

2008), and Human Givens therapy (HG; *k* = 1: Andrews et al., 2011). These studies were of variable quality, ranging from poor to good. Houghton and Saxon (2007) and McHugh et al. (2013) reported positive treatment outcomes following brief CBT of four and six sessions respectively. Despite the positive conclusions, both studies were of poor quality and suffered from significant methodological flaws. Moreover, there was disparity between studies in rates of improvement found.

Houghton and Saxon (2007) reported a 25% improvement rate among 44 patients experiencing anxiety who had complete data at pre- and post-assessment recorded at three months follow-up. Patients received, on average, 2.5 sessions of group psychoeducational CBT delivered by trained mental health nurses. The relatively low number of sessions attended by participants renders the acceptability of the intervention questionable; whilst the exclusive focus on patients experiencing anxiety and small sample size limits the generalisability of the findings. Due to post-data immediately following intervention being unavailable, it is also not possible to draw conclusions of the sustainability of treatment effects.

In contrast, McHugh et al. (2013) yielded an 84.6% RCSI rate for 13 patients scoring above clinical cut-off at intake following individual CBT delivered by graduate practitioners. Patients had common clinical presentations of depression and anxiety, and were eligible for treatment provided their difficulties were of mild to moderate severity. Of the initial 43 recruited patients, 19 completed therapy. Follow-up data assessed at 3-months post intervention, available for 8 patients, were indicative of maintenance in treatment effects. However, the initial sample size was small which, alongside a high level of attrition, introduced potential bias and limits generalisation.

Differences in the outcomes observed in the above CBT studies potentially reflect variations in treatment delivery formats, for instance the direct or indirect use of

specific therapeutic techniques, and the time elapsed between the completion of pre and post assessments.

Two primary care intervention studies considered the effectiveness of under-represented treatments. These studies benefitted from higher quality ratings and greater rates of intervention completion. In a study of the effectiveness of six-session PCT aimed at patients experiencing moderate to severe mental health difficulties, Gibbard and Hanley (2008) found 67.7% of patients having reliably improved and a large pre-post effect size of 1.2. These data represented outcomes of 697 patients (63%) with complete data, who commonly experienced depression or anxiety and were often treated concurrently with medication. Treatment was delivered by 12 counsellors whose theoretical orientation and training comprised PCT and who ranged in their clinical experience from students to accredited counsellors. Specific components and principles of PCT were poorly described and the effects of medication were not controlled.

Andrews et al. (2011) investigated the effectiveness of a HG approach delivered by three therapists. The HG approach was described as focusing on the patients' current situation, with an additional emphasis on past trauma and unmet emotional needs. Patients (n=124) presented with depression and/or anxiety, and over half were treated concurrently with medication. An overall pre-post effect size of 1.41 was found and an RCSI rate of 60.8% were reported for 79 patients (74.5%) whose pre-therapy scores fell within the clinical range. Subsequent analysis of treatment endings found that patients who had planned ends to treatment had better outcomes that those with unplanned termination.

All included studies drawn from specialist care services used pre-post designs to examine specific interventions encompassing Psychodynamic Interpersonal

Psychotherapy (PIT: Paley et al., 2008; Reuber et al., 2007) and Dialetical Behaviour Therapy (DBT: McFetridge & Coakes, 2010). These studies were of fair quality.

PIT interventions were described as emphasising the therapeutic relationship and conversation from which to focus on past and present interpersonal problems with a view to reduce symptoms. However, PIT studies differed in the problems of therapeutic focus, therapeutic durations, and treatment purity. Reuber et al. (2007) investigated the effectiveness of tailored PIT in reducing psychological distress and improving illness perceptions among a sample of 94 patients experiencing functional neurological symptoms. Treatment involved an average of 6 sessions in which an individualised and adapted PIT approach was delivered by one therapist. Findings indicated a significant improvement in post-treatment scores for 63 patients. Additionally, the authors reported improvement to be maintained at 6-months follow-up (n = 34). Therapeutic engagement was relatively low, with therapy non-completion observed for approximately 57.4% of patients, raising questions of the acceptability of treatment. Moreover, given that the therapy was tailored and a number of patients received supplementary CBT techniques, it is unclear as to how dilute or enhanced the treatment received was and, therefore, which approach is contributing to improvements.

Positive outcomes following PIT were also reported by Paley et al. (2008) among 67 patients described as experiencing enduring and severe mental health difficulties. Four therapists, trained in PIT, provided between 16 and 25 sessions to patients recruited consecutively from routine referral procedures. Fifty-seven patients (85.1%) completed therapy. Statistically significant reductions in distress were observed, with effect sizes of .76 and rates of RCSI of 34%. The greater rate of treatment completion observed by Paley et al. (2008), in comparison to Reuber et al. (2007), might suggest PIT is a more acceptable treatment of choice for patients referred

explicitly for mental health difficulties. The very nature of functional neurological symptoms potentially introduces different patient expectations of psychological therapies and their perceived benefits.

McFetridge and Coakes (2010) investigated the long-term effectiveness of a DBT informed approach for BPD. Participants were 40 women who had attended a therapeutic community setting and received between 8 to 12 months of individual and group therapy. Intervention was delivered by staff trained in the DBT approach, with weekly group supervision provided to support adherence. Approximately 5 years had elapsed between leaving the therapeutic programme and study participation. Post-therapy scores were significantly lower than pre-therapy scores, and those patients completing therapy were shown to have greater gains than those who withdrew. However, the sample represented only 34% of those attending the programme. Moreover, the immediate effects of treatment were not established introducing further limitations concerning the sustainability of improvement.

The final pre-post design study reviewed reported outcomes of an Emotional Support and Counselling (ESaC) intervention (Hodge et al., 2012) designed for patients (n = 35) with visual impairments and delivered in a voluntary sector setting. Treatment involved exploring patient emotional needs, validation of feelings, and counselling from a humanistic approach. Patients received 6 to 12 individual sessions delivered on a weekly basis by a qualified counsellor. Hodge et al. (2012) found significant improvements in psychological distress and stated patients demonstrating RCSI. However, the authors report only reliable improvement (54%) and the study was of poor quality due to a number of methodological weaknesses, such as under-reporting of patient characteristics, lack of consideration of confounding factors, and use of a small sample size.

The above studies demonstrated a degree of patient improvement following a range of interventions. However, observed outcomes across studies were highly variable. While such variability might be attributable to methodological differences, there remain unanswered questions concerning the impact of patient and contextual factors owing to: i.) the diversity in patient symptomology; ii.) theoretical underpinnings and change mechanisms of individual approaches and; iii.) delivery of interventions across varying lengths of treatment and by therapists of differing levels of experience. Furthermore, due to the nature of pre-post designs, it is not possible to determine whether outcomes are due to treatment specificity effects given the lack of intervention or control comparisons.

Only two studies, both of good quality, investigated comparative effects of treatments delivered in primary care (Holmqvist et al., 2014; Stiles et al., 2008), yielding conflicting findings. In a large sample (n = 5,613) UK study, Stiles et al. (2008) reported equivalent treatment outcomes across CBT, PCT, PDT, or these treatment approaches diluted/enhanced with one other treatment type. Holmqvist et al. (2014), in contrast, observed significantly less improvement in psychological distress among patients receiving supportive therapies than those in receipt of reflective (PDT) or directive approaches (CBT) in Swedish settings. Overall sample effect sizes were relatively similar; 1.39 (Stiles et al., 2008) and 1.37 (Holmqvist et al., 2014). However, RCSI rates were notably different. Holmqvist et al. (2014) found 34% of clinical patients achieving RCSI compared to 58.3% of patients in the Stiles et al. (2008) study. These differences potentially reflect variances in population and healthcare systems between countries and raise questions of transferability and comparability of outcomes.

Therapeutic Process Effects

Three studies were characterised by examining therapeutic processes associated with outcomes. Variability in outcomes as a function of treatment duration was the focus of two studies (Barkham et al., 2006; Stiles et al., 2015) and one study explored the impact of waiting times (Beck et al., 2014).

In primary care, Barkham et al. (2006) found treatment duration to be negatively associated with RCSI rates, with fewer sessions correlated with higher rates of RCSI. Full sample outcomes revealed a pre-post effect size of 1.51 and RCSI rate of 56.5%. While the study was generally of good quality, outcome completion rates were relatively low; with data representative of just 28.3% of the original dataset. More recently, a multi-sector study replicated the negative association between RCSI rates and treatment durations within primary care settings (Stiles et al., 2015). However, the same finding was not observed for pre-post effect sizes. Furthermore, differentiating by sector revealed similar rates of improvement regardless of treatment length in voluntary and workplace counselling sessions; suggesting importance in considering specific service effects and healthcare systems. Overall, a pre-post effect size of 1.89 and RCSI rate of 60% were found.

Average waiting times of 41.1 weeks between referral and first appointment were observed in patients attending a purposefully selected secondary care service (Beck et al., 2014). Waiting time for therapy was significantly associated with outcomes but not at a clinically meaningful level. An overall reliable improvement rate of 54.4% was reported. The representativeness of the sample was limited; the initial sample comprised only 23.5% of patients attending the service and was compounded by the pre-post return rate of 42.4%.

Therapist Effects

More recent focus of variability in treatment outcomes has concerned the role of therapists. Three studies considered therapist effects or factors associated with outcomes; two focused on therapist effectiveness (Armstrong, 2010; Saxon & Barkham, 2012) and one investigated therapist use of Rogerian therapeutic conditions (Murphy & Cramer, 2014).

In a voluntary sector study, Armstrong (2010) aimed to determine the effectiveness of minimally trained counsellors delivering brief interventions. The findings showed a pre-post effect size of .70 and a RCSI rate for all patients of 30.5%. Descriptive comparisons with primary care services suggested the participating counsellors were less effective (Armstrong, 2010). However, no formal analyses of therapist factors and their effect on patient outcomes were undertaken. Moreover, in primary care, Saxon and Barkham (2012) used advanced multilevel modelling techniques and found that factors at both patient and therapist levels, such as intake severity and risk caseload, contributed to poorer treatment outcomes. Therapists were found to account for 6.6% of variability in patient outcome, with below average therapists having fewer patients reaching RCSI. Saxon and Barkham (2010) found higher rates of RCSI (61.6%) and a larger effect size of 1.55.

In a student sample, Murphy and Cramer (2014) found that therapist and patient ratings of Rogerian therapeutic conditions were predictors of therapeutic progress and outcomes, whilst controlling for intake levels of patient distress. Change in scores was only explored across the first three sessions. By session three, a pre-post effect size of .85 was found and 24.2% of patients met RCSI criteria. Reliable improvement was found for 22.5% of patients. The sample size was small and the findings do not

evidence lasting change. Nonetheless, the results reflect the importance of the therapeutic relationship in patient outcomes.

Summary of Outcomes

Table 4 presents effect sizes for the 15 studies reporting or providing data from which effect sizes could be estimated. Effect sizes were large across all studies except Reuber et al. (2007) who reported a medium pre-post change. The average effect size was 1.23 (95% CI: 1.0-1.5, n = 15). Primary care studies showed larger effect sizes than the other service provisions, with specialist and alternative providers demonstrating somewhat smaller effects.

Of the 22 studies, 18 reported estimates of RCSI and/or associated change categories. Table 5 shows the proportion of patients meeting change categories provided by each study. Similar to the above, primary care studies tended to have higher rates of RCSI compared to other settings. Additionally, greater proportions of patients meeting reliable deterioration criteria were found within secondary, specialist, and alternative care providers.

Table 4.

Reported and calculated study effect sizes

Lead Author	Effect Size (d)
Primary Care	
Andrews (2011)	1.41
Barkham (2006)	1.51
Barkham (2012)	1.95
Evans (2002)	1.44*
Gibbard (2008)	1.20
Holmqvist (2014)	1.37
Mullin (2006)	1.43*
Saxon (2012)	1.55
Stiles (2008)	1.39
Secondary Care	
Barkham (2001)	.87
Specialist Services	
Paley (2008)	.76
Reuber (2007)	.42*
Alternative Provider	
Armstrong (2010)	.70
Connell (2008)	1.57
Murphy (2014)	.85

* Calculated from available data

Table 5.

Proportions of study patients meeting change criteria

Lead Author	RCSI (%)	Reliable Improvement (%)	No reliable change (%)	Reliable Deterioration (%)
Primary Care				
Andrews (2011)	60.8	13.9	-	-
Barkham (2006)	71.7	11.6	15.6	1.1
Barkham (2012)	58.3	71.7		
Evans (2002)	49.0	25.0	26.0	1.0
Evans (2003)	-	77.8	-	-
Gibbard (2008)	-	67.7	30.9	1.4
Holmqvist (2014)	34.0	43.0	-	-
Houghton (2007)	25.0	20.5	45.0	.09
McHugh (2013)	84.6	50.0	-	-
Mullin (2006)	55.0	19.0	26.0	1.0
Saxon (2012)	61.6	-	-	-
Stiles (2008)	58.3	19.4	21.1	1.2
Secondary Care				
Barkham (2001)	39.0	15.0	40.0	6.0
Beck (2014)	-	54.4	34.0	2.9
Specialist Care				
Paley (2008)	40.0	18.0	37.0	5.0
Alternative Providers				
Armstrong (2010)	30.5	17.8	44.0	7.6
Connell (2008)	54.3	21.8	23.2	.70
Murphy (2014)	24.2	22.5	43.5	3.2

Discussion

Uniquely focused on the CORE-OM, the current review aimed to critically evaluate PBE studies and summarise patient outcomes according to mental health provider. Twenty-two studies satisfied inclusion criteria. Within these studies, the CORE-OM was used for a range of evaluative purposes comprising four main areas: global service effects, specific intervention and comparative intervention effects, and contextual effects of therapeutic processes and therapist factors. The CORE-OM was

most frequently used as a means of evaluating the outcomes of specific therapeutic interventions. Contextual factors, as they impact on outcomes, were less commonly investigated.

Outcomes derived across studies were generally positive and suggested a range of mental health providers were effective in reducing patient psychological distress. The average uncontrolled effect size was large (d = 1.23) and broadly consistent with the Cahill et al. (2010) finding of 1.29 for CMHPs. In contrast, the average rate of RCSI (41.5%) was lower than the 56% reported by Cahill et al. (2010). Differences in obtained RCSI rates might be attributable to the current value reflecting outcomes of a broader range of service providers and clinical presentations, with the Cahill et al. study finding restricted to primary care settings and CMHPs. However, in keeping with such reporting, the average RCSI rate for the present primary care studies was equivalent (55.8%); with seven such studies meeting the 54% RCSI benchmark used by Cahill et al. (2010).

In terms of outcomes among the remaining care settings, due to a lack of studies available and/or providing data, it was not possible to determine average effect sizes or RCSI rates. However, tentative descriptive trends were able to be discerned. Findings showed smaller effect sizes and lower rates of RCSI among secondary, specialist, and alternative care providers. The smallest magnitudes of change were observed among specialist and voluntary services, with greater proportions of patients meeting criteria for reliable deterioration across services beyond primary care. Only one university counselling study yielded outcomes consistent with those of primary care.

While it should be noted that treatment effects generally remained moderate to large within these studies, the occurrence of such descriptive patterns warrants further consideration. The limited number of studies examining direct influences on outcomes,

however, reduces any such exploration to tentative suggestions. For instance, three studies (Barkham et al., 2001; Mullin et al., 2006; Saxon & Barkham, 2012) found that greater initial severity was associated with poorer outcomes. Thus, patients attending secondary and specialist care services might be expected to have greater levels of psychological distress and, therefore, poorer outcomes. Available profile data offer some support for such assertions. Secondary care patients averaged somewhat higher intake scores than primary care patients, as did those attending specialist provisions with the exception of Reuber et al.'s study. Moreover, patients studied within specialist service tended to have axis-II disorders or associated presentations; suggesting patient complexity might be influential to outcomes.

However, initial intake severity is not sufficient in itself to explain between service outcome variability, particularly in the context of patients within Reuber et al.'s study and alternative providers having severity scores in keeping with patients presenting to primary care. These findings suggest outcomes are influenced by alternate factors. One possibility, remaining at the patient level, is that variability occurs as a result of the idiosyncrasies of specific patient subsets. A further possibility, and somewhat evidenced within the review findings, are the effects of service delivery/design components and therapists.

In terms of service delivery, preliminary evidence suggests one factor to be considered as impacting outcomes is that of treatment duration (Barkham et al., 2006; Stiles et al., 2015). The latter study highlighted that the influence of duration is likely to be dependent on the type of service, which emphasises the need for further understanding of how service providers are differentiated in respect of both their attending patients and treatment formats. Due to inadequate numbers of comparative

studies or groups, it was not possible to determine distinct patterns of influence regarding specific treatment modalities.

A final consideration concerns therapist factors and effects. The three available studies (Armstrong, 2010; Murphy & Cramer, 2014; Saxon & Barkham, 2012) indicated therapist factors concerning the therapeutic relationship, level of training, and caseload variables were influential to outcome. However, there remains insufficient evidence from which to determine whether services and their outcomes are differentiated by therapist factors.

Quality of Evidence

The strength of evidence was informed by assessment of study methodological quality using an adapted version of the Downs and Black (1998) checklist. The current assessment found an absence of studies of the highest quality and identified similar quality issues as reported by Cahill et al. (2010); namely that PBE studies suffered from high levels of potential sources of internal bias. Within the present review, issues of bias primarily arose due to: i.) a lack of representativeness due to small samples, rates of attrition, exclusive focus on particular subsets of patients, or under-reporting of patient characteristics; ii.) inadequate discussion or use of appropriate statistical analyses taking account of potential confounding factors; and iii.) under-reporting of therapist characteristics, intervention compliance, and adherence. These issues render generalisability of the findings questionable and suggest caution in their interpretation, with the potential of outcomes being over-estimated.

Review Limitations

The review was dominated by primary care studies and identified a scarcity of PBE studies using the CORE-OM beyond these settings. While such finding is

plausible in the context of the CORE-OM having been initially established for use in primary care and only more recently validated across other settings (e.g., Barkham et al., 2005; Connell, Barkham, & Mellor-Clark, 2007), it suggests caution in the interpretation of the overall service effects currently yielded.

In addition, the number of studies obtained reflects the review search strategy and inclusion criteria. For instance, the extent to which the CORE-OM has been used within PBE research was potentially underestimated by the following: i.) the limited number of databases searched (n = 3); ii.) the lack of forward citation searching, particularly concerning the original publication of the CORE-OM (Barkham et al., 1998); and iii.) the requirement of studies to have used the CORE-OM as the primary outcome tool. Moreover, restricted inclusion to peer reviewed and English language papers renders the outcomes of the current review subject to publication bias' and, therefore, suggests further caution is warranted in the interpretation of findings.

The degree of heterogeneity between studies prevented formal meta-analyses from being undertaken. As such, studies could not be synthesised beyond their focus of service delivery components and descriptive analyses. Without systematic pooling, confounding factors could not be adequately controlled introducing bias into the overall treatment effect observed.

Clinical Implications

A range of mental health providers appear to be effective in their treatment of psychological distress. Such findings provide evidence of effectiveness from routine practice from which to complement and begin to redress the balance of efficacy based studies. The present review also highlighted that the comparative effectiveness of mental health services has not been directly addressed. Direct service comparisons have potentially important implications for policy initiatives concerning improving access to

psychological treatments and quality of care through highlighting the role of specific providers and their potential utility in the provision of mental healthcare to be assessed.

Future Research

In consideration of the review findings, there are key gaps in knowledge and research concerning outcomes and service effectiveness that require address. Of the CORE-OM research completed, a clear pattern emerged indicating that intervention effects have been the most widely examined. However, these studies generally lacked robustness and tended to be of weaker methodological design. As such, reliance on these outcomes to suggest effectiveness of services is questionable. Moreover, the majority of intervention studies failed to provide, or adequately report, follow-up data; thereby inhibiting development of PBE knowledge base concerning the maintenance of therapeutic gains. There is a need, therefore, for intervention studies to improve their quality; with particular attention to their low levels of reporting and lack of control of confounds alongside the development of longitudinal PBE designs.

Current findings remain confronted by issues of generalisability and transferability owing to the dominance of primary care studies. Further research is required to demonstrate the effectiveness of under-represented services, such as secondary or alternative providers of mental health care. However, it cannot be assumed that services are accessed by the same types of patients or provide the same types of treatments delivered under the same conditions. As such, there is an additional need to refine the exploration and reporting of organisations' individual structures, service design factors, and therapist and patient characteristics. In doing so, greater understanding can be gained of the types of services, and their processes, able to meet specific patient needs. Alternately stated, further knowledge of key patient, therapist,

and service characteristics can help inform practitioners of who is most likely to respond and engage with a given provider.

Moreover, PBE studies have largely focused on a single aspect of service delivery rather than attempting to provide more comprehensive understandings of the complex interactions between multiple levels of therapeutic processes and outcomes. Greater identification and synergy of pivotal and service specific factors affecting outcomes is needed in order to accurately determine the effectiveness of services.

Conclusion

PBE studies using the CORE-OM suggested a range of mental health providers are effective in reducing psychological distress. Treatment outcomes across mental health services averaged an uncontrolled effect size of 1.23 and RCSI rate of 41.5%. However, multiple factors appear to contribute toward outcomes and there is a need for greater synergy of such factors in order that effectiveness be fully and accurately determined, alongside increased development of the evidence base for services beyond primary care.

References

*denotes studies included in the review

* Andrews, W., Twigg, E., Minami, T., & Johnson, G. (2011). Piloting a practice research network: A 12-month evaluation of the Human Givens approach in primary care at a general medical practice. *Psychology and Psychotherapy: Theory, Research and Practice*, *84*, 389-405. doi: 10.1111/j.2044-8341.2010.02004.x

*Armstrong, J. (2010). How effective are minimally trained/experienced volunteer mental health counsellors? Evaluation of CORE outcome data. *Counselling and Psychotherapy Research*, *10*, 22-31. doi: 10.1080/14733140903163284

Barkham, M., Evans, C., Margison, F., & McGrath, G. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health*, *7*, 35. doi: 10.1080/09638239818328

*Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*, 184.

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *The British Journal of Psychiatry*,*186*, 239-246. 10.1192/bjp.186.3.239

*Barkham, M., Connell, J., Stiles, W. B., Miles, J. N., Margison, F., Evans, C., & Mellor-Clark, J. (2006). Dose-effect relations and responsive regulation of treatment duration: the good enough level. *Journal of consulting and clinical psychology*, *74*, 160. doi: 10.1037/0022-006X.74.1.160

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence*. Chichester: Wiley.

*Barkham, M., Stiles, W. B., Connell, J., & Mellor-Clark, J. (2012). Psychological treatment outcomes in routine NHS services: what do we mean by treatment effectiveness?. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*, 1-16. doi: 10.1111/j.2044-8341.2011.02019.x

*Beck, A., Burdett, M., & Lewis, H. (2015). The association between waiting for psychological therapy and therapy outcomes as measured by the CORE-OM. *British Journal of Clinical Psychology*, *54*, 233. doi: 10.1111/bjc.12072

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology*, *49*, 421-453. doi: 10.1348/014466509X470789

*Connell, J., Barkham, M., & Mellor-Clark, J. (2008). The effectiveness of UK student counselling services: an analysis using the CORE System.*British Journal of Guidance & Counselling*, *36*, 1-18. doi: 10.1080/03069880701715655

Department of Health. (2004). *Standards for better health*. London: DoH. Retrieved from:http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4132991.pdf

Department of Health. (2012). *No health without mental health: Implementation Framework*. London: DoH. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216870/No-Health-Without-Mental-Health-Implementation-Framework-Report-accessible-version.pdf

Department of Health. (2014). *Closing the gap: Priorities for essential change in mental health*. London: DoH. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/281250/Closing_the_gap_V2_-_17_Feb_2014.pdf

*Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM. *The British Journal of Psychiatry*, *180*, 51-60. doi: 10.1192/bjp.180.1.51

*Evans, C., Connell, J., Barkham, M., Marshall, C., & Mellor-Clark, J. (2003). Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy*, *10*, 374-388. doi: 10.1002/cpp.384

*Gibbard, I., & Hanley, T. (2008). A five-year evaluation of the effectiveness of person-centred counselling in routine clinical practice in primary care. *Counselling and Psychotherapy Research*, *8*, 215-222. doi: 10.1080/14733140802305440

*Hodge, S., Barr, W., Bowen, L., Leeven, M., & Knox, P. (2013). Exploring the role of an emotional support and counselling service for people with visual impairments. *British Journal of Visual Impairment*, *31*, 5-19. doi: 10.1177/0264619612465168

*Holmqvist, R., Ström, T., & Foldemo, A. (2014). The effects of psychological treatment in primary care in Sweden—A practice-based study. *Nordic journal of psychiatry*, *68*, 204-212. doi: 10.3109/08039488.2013.797023

*Houghton, S., & Saxon, D. (2007). An evaluation of large group CBT psycho-education for anxiety disorders delivered in routine practice. *Patient education and counseling*, *68*, 107-110. doi: 10.1016/j.pec.2007.05.010

Jacobs, R. (2009). *Investigating Patient Outcome Measures in Mental Health. CHE Research Paper Number 48*. York: Centre for Health Economics, University of York.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*, 363-374. doi: 10.2307/2529786

*McFetridge, M., & Coakes, J. (2010). The Longer-term Clinical Outcomes of a DBT-informed Residential Therapeutic Community; An Evaluation and Reunion. *Therapeutic Communities*, *31*, 406-416.

*McHugh, P., Brennan, J., Galligan, N., McGonagle, C., & Byrne, M. (2013). Evaluation of a primary care adult mental health service: Year 2. *Mental health in family medicine*, *10*, 53-59.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*, 264-269. doi:10.1371

*Mullin, T., Barkham, M., Mothersole, G., Bewick, B. M., & Kinder, A. (2006). Recovery and improvement benchmarks for counselling and the psychological

therapies in routine primary care. *Counselling and Psychotherapy Research*, *6*, 68-80. doi: 10.1080/14733140600581515

*Murphy, D., & Cramer, D. (2014). Mutuality of Rogers's therapeutic conditions and treatment progress in the first three psychotherapy sessions. *Psychotherapy Research*, *24*, 651-661. doi: 10.1080/10503307.2013.874051

*Paley, G., Cahill, J., Barkham, M., Shapiro, D., Jones, J., Patrick, S., & Reid, E. (2008). The effectiveness of psychodynamic-interpersonal therapy (PIT) in routine clinical practice: A benchmarking comparison. *Psychology and psychotherapy: Theory, research and practice*, *81*, 157-175. doi: 10.1348/147608307X270889

*Reuber, M., Burness, C., Howlett, S., Brazier, J., & Grünewald, R. (2007). Tailored psychotherapy for patients with functional neurological symptoms: a pilot study. *Journal of psychosomatic research*, *63*, 625-632.

Samoocha, D., Bruinvels, D. J., Elbers, N. A., Anema, J. R., & van der Beek, A. J. (2010). Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. *Journal of medical Internet research*, *12*, e23. doi: 10.2196/jmir.1286.

*Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, *80*, 535. doi: doi: 10.1037/a0028898

Seligman, M. E. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American psychologist*, *50*, 965 - 974.

Shadish, W. R., Matt, G. E., Navarro, A. M., Siegle, G., Crit-Christoph, P., Hazelrigg, P., et al. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, 65, 355–365.

Shadish, W. R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*,126, 512–529. doi: 10.1037/0033-2909.126.4.512

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American psychologist*, 752 – 760. doi: 10.1037/0003-066X.32.9.752

Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological medicine, 36*, 555-566. doi: 10.1017/S0033291706007136

*Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: replication in a larger sample. *Psychological medicine, 38*, 677-688.

*Stiles, W. B., Barkham, M., & Wheeler, S. (2015). Duration of psychological therapy: relation to recovery and improvement rates in UK routine practice. *The British Journal of Psychiatry, 207*, 115-122. doi: 10.1192/bjp.bp.114.145565

Appendix A: Adapted Quality Checklist (Cahill et al., 2010)

<copyright information removed>

Appendix A: Continued

<copyright information removed>

Appendix A: Continued

<copyright information removed>

Appendix A: Continued

<copyright information removed>

Appendix B. Individual Item Assessment of Methodological Quality

Table 3.

Individual Item Assessment of Methodological Quality per Study

	Reporting											External Validity												Internal Reliability					Internal Validity					Total
Item	1	2	3	4	5	6	7	8	9	10	11	12 a	12 b	13	14 a	14 b	15 a	15 b	15 c	16	17	18	19	20	21	22	23	24	25	26	27	28	
Lead Author																																	
Primary Care																																	
Andrews 2011	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	0	0	0	1	1	1	24
Barkham 2006	1	1	1	1	0	1	1	0	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	0	0	1	20
Barkham 2012	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	21
Evans 2002	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	1	1	0	1	0	1	0	0	0	0	15
Evans 2003	1	1	1	0	1	1	1	0	0	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1	19
Gibbard 2008	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	0	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	19
Holmqvist 2014	1	1	1	0	1	1	1	0	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	0	1	0	1	1	0	0	0	20
Houghton 2007	1	1	0	1	0	1	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	1	1	0	1	0	1	0	0	0	0	14
McHugh 2013	1	1	1	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	12
Mullin 2006	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1	20
Saxon 2012	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	0	1	0	0	1	1	0	1	22
Stiles 2008	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	0	1	1	0	1	23
Secondary Care																																	
Barkham 2001	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	20
Beck 2015	1	1	1	0	1	1	1	0	0	1	1	0	1	0	1	0	0	1	1	1	0	1	1	1	0	1	0	1	1	1	0	0	20

	Reporting											External Validity											Internal Reliability					Internal Validity					Total
Item	1	2	3	4	5	6	7	8	9	10	11	12		13	14		15			16	17	18	19	20	21	22	23	24	25	26	27	28	
												a	b		a	b	a	b	c														
Lead Author																																	
Specialist Care																																	
McFetridge 2010	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0	1	1	1	0	1	0	1	1	0	0	0	16
Paley 2008	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	19
Reuber 2007	1	1	1	1	0	1	1	0	0	1	1	0	1	0	1	0	0	1	1	0	0	1	1	1	0	1	0	1	0	0	0	0	17
Alternative Provider																																	
Armstrong 2010	1	1	1	0	0	1	1	0	0	1	1	0	1	0	1	1	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	18
Connell 2008	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	1	0	1	0	0	1	1	0	1	22
Hodge 2012	1	1	1	1	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	14
Murphy 2014	1	1	1	1	0	1	1	0	0	1	1	0	1	0	1	0	0	0	1	1	1	1	1	1	0	1	0	0	0	1	1	0	19
Stiles 2015	1	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	1	0	1	23

Note: 1 = yes, 0 = no or unable to determine

Empirical Report

The effectiveness of Voluntary Sector Organisations in comparison with the National Health Service: Patient profiles, therapist effects & treatment outcomes

Abstract

Objectives: Primary study aims comprised: i.) compiling profiles of patients, service delivery components, and therapists among voluntary sector organisations (VSOs); ii.) comparing VSO profiles to that of a National Health Service (NHS) comparator sample; iii.) determine the effectiveness of VSOs; and iv.) apply multilevel modelling techniques to determine predictors of outcomes and size of therapist effects among VSOs.

Design: A quantitative cohort design was used.

Method: Secondary analyses of a large practice-based dataset were undertaken. The sample comprised 2,157 patients and 50 therapists from voluntary sector organisations (VSOs), and a further 7,985 patients and 83 therapists from the NHS. Included patients represented treatment completers who had valid pre-post CORE-OM data available. Descriptive analyses were used to compile profile information and multilevel modelling procedures employed to determine predictors of therapeutic outcomes and therapist effects.

Results: Patient severity profiles were broadly similar between VSO and NHS samples. VSOs and NHS providers were primarily differentiated by their therapeutic orientations, treatment durations, and therapist caseload sizes. Outcomes between sectors were broadly comparable, with recovery rates of 58% and 61% for VSO and NHS providers respectively. VSOs had therapist effects of 4.5%, considerably smaller than therapist effects among the NHS sector (12.7%).

Conclusion: VSOs are equivalent in their effectiveness at reducing psychological distress in their patients.

Practitioner Points

- VSOs might offer a useful alternative to NHS mental health provisions from which to meet demand for psychological therapy
- Key predictors of VSO outcomes concerned patient sociodemographic variables and number of therapies provided
- Further research is recommended to understand why people attended and use VSOs
- Further research is required into the characteristics and skill sets of VSO therapists

Limitations

- Data represented a completer only sample
- Therapist factors were limited to those able to be aggregated from existing variables

Introduction

Recent policy initiatives' focus on mental health provision has concerned improving access to psychological therapies and the quality of treatment delivered (Department of Health [DoH], 2011; 2012). The impetus to address such issues stemmed from a growing demand for psychological treatments, resulting in a large unmet need for mental health care and inequity in access to a full range of interventions, alongside the identification of poor care and outcomes (Care Quality Commission [CQC], 2015; National Health Service [NHS] England, 2015). Key objectives, aimed to improve quality, outlined by the DoH (2012) encompassed providers: i.) demonstrating their effectiveness; ii.) integrating community and person-centred care; and iii.) adopting a recovery-orientated approach from which to support patients in the management of their own care.

Additionally, the *Any Qualified Provider* (AQP) framework, proposed plans to extend patient choice and commissioning processes to mental health providers who could meet rigorous quality standards (DoH, 2011). The intention was to promote treatment engagement by offering patients greater choice and control of their care in conjunction with widening access to treatments and affording an arena for new providers to address gaps in current provisions. Identified within the initiative was the potential of Voluntary Sector Organisations (VSOs) to make a significant contribution to, and be partners in, the delivery of mental health care.

Approximately 35,000 VSOs provide health and social care support; ranging from counselling, advocacy, and advice to employment and community-based inclusion schemes, each aimed at promoting the mental and physical well-being of service-users (Curry, Mundle, Shiel, & Weaks, 2011). In addition, VSOs often undertake key roles of engaging communities and facilitating access to services through campaigning, service

user involvement, and the development of new and innovative approaches to care (e.g., Appleby, 2009; Curry et al., 2011; Kendal & Knapp, 1999; Miller, 2013;). It is such practices that have led VSOs to be attributed with characteristics of responsiveness and trustworthiness, and as offering inclusive and personalised support believed to demarcate them as unique providers of mental health care with potential to add value to current practice (Buckingham, 2009; DoH, 2007).

However, evidence of VSO characteristics, such as those described above, is primarily anecdotal. The DoH (2006) identified a number of areas requiring address in order that involvement of VSOs in the provision of mental health care is improved. These areas included gaining a greater understanding of the remit of VSOs and evidencing of VSO accountability and fitness for purpose. Despite requirements having been identified, the impact of VSO practice on patient outcomes remains lacking, with many reviews concluding that the evidence-base for VSOs does not yet exist (e.g. McMillan, 2011; Miller, 2013).

Therapist Effects

Within the wider literature, evidence for the use of psychotherapy in the treatment of mental health disorders has been well established (e.g., Roth & Fonagy, 2006). More recent focus has concerned examining determinants of therapeutic change and understanding observed outcome variability. Particular emphasis has been placed on the differential effects of individual therapists (i.e., therapist effects) on patient outcomes (e.g., Crits-Christoph & Mintz, 1991; Lutz, Leon, Martinovich, Lyons, & Stiles, 2007; Martindale, 1978). In the context of quality assurance frameworks, understanding such influences are crucial to accurately determining the effectiveness of services and improving patient care (e.g., Kim, Wampold, & Bolt, 2006; Saxon & Barkham, 2012).

Therapist effects have been determined in both clinical trials research and routine clinical practice (Baldwin & Imel, 2013; Brown, Lambert, Jones, & Minami, 2005; Green, Barkham, Kellett, & Saxon, 2014). Resultant therapist effects have varied substantially; from no effect to 50% (e.g., Elkin, Falconnier, Martinovich, & Mahoney, 2006; Crits-Christoph & Mintz, 1991). However, the most frequently reported effects lie between 5 and 10% (Crits-Christoph & Mintz, 1991).

A number of factors have been found to influence therapist effects that might account for the degree of observed variation between studies. Greater therapist effects have been associated with increased patient intake severity (e,g,. Kim et al., 2006; Saxon & Barkham, 2012), larger therapist caseloads (e.g., Vocisano et al., 2004), and higher than average therapist risk-caseloads (Saxon & Barkham, 2012). In contrast, smaller therapist effects have been attributed to manualised therapies (Crits-Christoph & Mintz, 1991) and higher levels of therapist training and experience (Huppert et al., 2001).

Further Determinants of Outcomes

In addition to therapist effects, Lambert (1992) maintained that therapeutic outcomes also vary in accordance with patient and treatment related factors. Patient variables tend to fall within three main categories of exploration: sociodemographic factors, clinical presentation, and latent factors. Patient sociodemographic factors, whilst extensively examined, have yielded few consistent findings in their relation to outcome. Where significant effects of sociodemographic variables have been observed, these have shown a tendency toward patients who are unemployed, of an ethnic minority, or living alone benefitting least from therapy than their counterparts (e.g., Hamilton & Dobson, 2002; Lam & Sue, 2001; Ostler et al., 2001). A more frequently observed finding concerns patient intake severity, greater levels of patient severity have

been associated with poorer outcomes (Garfield, 1994; Saxon & Barkham, 2012). Latent patient characteristics, such as motivation, ego-strength, and expectations, have demonstrated relatively small contributions to outcomes (e.g., McBride et al., 2010).

Investigations of process factors, such as those relating to the therapeutic alliance, specific treatments, and treatment delivery, have consistently shown that the therapeutic alliance is a strong predictor of outcome (e.g., Arnow et al., 2013). Data from routine practice, alongside a series of meta-analytic studies, have suggested that different therapeutic approaches produce equivalent outcomes whether delivered in pure form or combined with one other approach (e.g., Stiles, Barkham, Mellor-Clark, & Connell, 2007). These findings suggest that intervention characteristics are not sufficient in themselves to account for variability in outcomes. However, more consistent evidence has emerged concerning treatment duration, in which shorter treatment lengths are associated with better outcomes (Barkham et al., 2006).

Multilevel Modelling

Given that factors identified as being associated with psychotherapeutic outcomes encompass different levels of data structure, data analytic techniques that can account for variability at these different levels and determine their relative effects on outcomes are required. Multilevel modelling (MLM) offers such an approach. MLM assumes data exist within such hierarchical structures, in which data at one level (e.g., patients) are nested within observations at another level (e.g., therapists). Each level is a potential source of variability in the dependent variable of interest (e.g., outcome score). Traditional means of statistical analyses allow effects to be determined at one level, either level one or level two, which can inflate the risk of Type I error (Kim et al., 2006). In contrast, MLM affords the explicit modelling of variance at each level of the data structure from which estimates can be derived (Raudenbush & Bryk, 2002). MLM

has been advocated as a means of determining therapist effects (Wampold & Brown, 2005).

Current Study

The above research indicates key gaps in knowledge from which to determine the value of VSOs as partners in mental healthcare delivery. If VSOs are to play a pivotal role, key areas concerning a comprehensive understanding of VSO typology, models of care, and patients are required alongside determining both their individual effectiveness and comparative effects with other providers. In order that these issues are accurately addressed, VSO outcome data needs to be subject to the same rigorous methodological procedures (e.g., MLM) as data drawn from NHS providers.

As such, the present study aimed to:

i.) establish the effectiveness of voluntary sector services compared to NHS services, with consideration of the contributions of, and differences in, factors concerning client presenting profiles, therapist effects, and service delivery

ii.) determine which of the factors broadly outlined above contribute to effective outcomes in voluntary services.

The specific study objectives were as follows:

i.) Compile profiles of VSO patients and their presentations together with service delivery components (e.g., type and duration of interventions)

ii.) Compare derived VSO and NHS profiles

iii.) Apply multilevel modelling approach to investigate how the variables concerning client characteristics, types of treatment, dose effect, and therapist effects contribute to patient outcomes within the sample of voluntary sector services.

iv.) Determine the size and variability of therapist effects in the voluntary sector sample as compared with NHS comparator.

Method

Design

A quantitative cohort design was used, involving the secondary analyses of a large practice-based dataset comprising routinely collected data from multiple NHS and non-NHS mental health providers. The focus of the secondary analyses was to contrast voluntary and NHS sectors on a range of variables, and to employ MLM techniques to determine predictors of patient outcomes and extent of therapist effects.

Original Dataset

The original dataset, the CORE National Research Database-2011, comprised data from 104,474 patients (66.2% female; mean age = 35.88, SD = 13.28) and 2,442 therapists who saw between 1 and 1,658 patients each (M = 360.92, SD = 362.67). Patients were referred to, or seeking treatment from, UK mental health or counselling provisions between April 1999 and November 2011. In total, data were drawn from 52 sites nationally. Services contributing to the dataset included seven primary care services (27,012 patients), eight secondary care services (7,348 patients), two tertiary care services (4,460 patients), fifteen voluntary sector services (21,804 patients), two private practices (442 patients), eight workplace counselling services (17,204 patients), and ten university counselling centres (26,198 patients). Patients were allocated to treatment according to service usual procedures.

Study Specific Dataset

Development of the study dataset comprised the use of a series of syntax commands (SPSS version 21.0; IBM Corp., 2012). These commands were used to perform tasks concerning data cleansing and formatting, calculations and aggregations, and to develop numerical codes for categorical variables. A series of flag variables were created to simplify identification of cases where they had missing data on one or more variables under consideration. Next, successive filters were applied according to the study inclusion criteria described below. Inclusion criteria were applied at both patient and therapist levels of data.

Patient Selection. For the purposes of the present study, patients were selected from primary, secondary, and tertiary NHS sites, and voluntary sector services, and required to meet the following criteria: i.) received individual, one-to-one therapy, ii.) adult patients aged between 16 and 95 years, iii.) received two or more sessions, comprising an assessment session and treatment, iv.) had a planned end to treatment as described by their therapist, v.) returned data on all variables under consideration, and vi.) returned valid pre and post CORE-OM forms.

Within the original dataset, a number of patients had more than one episode of treatment recorded. To ensure patients were only included once the first recorded episode was selected. Patients were excluded who were in receipt of group or couples therapy, including those receiving both individual and group therapy, or where therapy type was missing (21,137 excluded). The purpose of focusing on individual therapy was due to the potential confounds of group therapy in determining therapist effects owing to the contribution from multiple individuals and to be in keeping with prior research on therapist effects for comparison. Of the remaining 33,172 patients, 488 were excluded due to missing data on age or falling outside of the specified age range

(16 - 95 years), 3,704 were excluded because they did not attend or received an assessment only session, 7,951 were excluded having had an unplanned ending to treatment or type of ending not indicated, and 3,446 had missing data on demographic and process variables under study. In total, 14,056 patients returned valid pre and post-treatment forms: 998 patients did not return valid pre or post-treatment forms; 2,379 returned pre-treatment forms only; and 150 returned post-treatment forms only.

Therapist Selection. Further variables, at the therapist level, were derived from the existing data. Such variables were used to improve the quality of the data by giving consideration to potential selective reporting and the minimum unit recommendations required for MLM procedures.

Prior research using primary care datasets has addressed case selection bias through employing a therapist return rate criterion (e.g., Saxon & Barkham, 2012). The criterion was based on return rate targets, currently 90 per cent or more, set by the Improving Access to Psychological Therapies (IAPT) programme (Department of Health [DoH], Mental Health Programme, 2008). However, there is a lack of published guidance on the data collection performances of services beyond primary care and it is unclear as to how representative the 90 per cent return rate would be for such services. Not wishing to ignore the issue of selective reporting, and finding a pre-post measure completion rate ranging between 4 and 100 per cent among therapists, a 'good enough' return of 50 per cent or above was adopted. Furthermore, in order to satisfy the minimum recommendations for MLM procedures outlined by Hox (1998), therapists were required to have worked with a minimum of 20 eligible participants.

Adopting a minimum return rate of 50 per cent yielded 13,666 patients. Of these, 10,142 patients were treated by a therapist with a caseload of 20 or more patients resulting in a sample of 133 therapists. The therapist sample was in keeping with the

level 2 unit requirements (minimum 50 units) for multilevel modelling procedures outlined by Maas and Hox (2004).

Participants

The final dataset included 10,142 patients derived from five primary care NHS sites (n = 7,369), five secondary care NHS sites (n = 587), one tertiary NHS site (n = 29) and eight voluntary sector services (n = 2,157). Sample characteristics for included and excluded patients of the study sample are summarised in Table 1.

Table 1.

Summary characteristics of included and excluded patients

Characteristic	**Included (M, SD)** **n = 10,142**	**Excluded (M, SD)** **n = 50,488**	**t (d.f.)**
Age	41.5 (13.1)	38.3 (13.0)	-22.4 (14750.5)*
Intake score	18.3 (6.3)	18.8 (7.0)	6.3 (16506.6)*
Sessions Attended	8.8 (10.2)	10.6 (16.5)	12.7 (29634.5)*
	Included (%)	Excluded (%)	Chi-Squared (d.f.)
Female	69.9	63.6	144.4 (1)*
White-European	94.8	86.6	527.8 (1)*
Employed	73.8	68.8	97.6 (1)*
Living with Other	75.9	71.8	68.3 (1)*

* $p \leq .001$

Comparisons between patients included and excluded from the final sample revealed included patients were significantly older ($p \leq .001$) and attended fewer sessions ($p \leq .001$). Excluded patients had significantly higher pre CORE-OM scores (p

≤ .001). The included sample were more likely to be female, of White/European ethnicity, in employment, and living with an other (all $p \leq .001$).

Measures

Outcome Measure. Patients completed the Clinical Outcomes for Routine Evaluation-Outcome Measure (CORE-OM; Barkham et al., 1998, 2005; Evans et al. 2002; Appendix A) prior to therapy and at their final session. The CORE-OM is a 34-item self-report inventory of psychological distress. The measure encompasses domains of subjective well-being, symptoms (anxiety, depression, physical problems, and trauma), functioning (general functioning, close relationships, and social relationships) and risk (risk to self and risk to others). The latter domain comprises items addressing both risk to self and risk to others. Each item is scored on a 5-point scale, anchored 0 '*not at all'* to 4 '*most or all of the time*'. Clinical scores are derived by calculating the mean of completed items multiplied by 10; yielding scores between 0 and 40. Higher scores are indicative of greater levels of distress. Recommended cut-off scores have been established: <10 low level (non-clinical), 10-14 mild, 15-19 moderate, 20-24 moderate-severe, and ≥25 severe. The CORE-OM has established psychometric properties, with an internal consistency of .94 (Barkham et al., 2001) and test-retest correlations of ≥.80 (Barkham, Mullin, Leach, Stiles, & Lucock, 2007).

Assessment and End of Therapy Forms. Contextual information was gained from therapists using the Clinical Outcomes in Routine Evaluation Therapy Assessment Form (CORE-TA; Mellor-Clark & Barkham, 2006; Appendix B) and the Clinical Outcomes in Routine Evaluation End of Therapy Form (CORE-ET; Mellor-Clark & Barkham, 2006; Appendix C). At intake, therapists completed the CORE-TA, which provided information concerning referral, patient demographic characteristics, concurrent treatment with medication, patient risk, and patient presenting problem(s) -

including details of their severity and duration. Fourteen categories of presenting problems are considered: depression, anxiety, psychosis, personality problems, cognitive/learning difficulties, eating disorder, physical problems, addictions, trauma/abuse, bereavement, self-esteem, interpersonal problems, living/welfare and work/academic. Recorded on the CORE-ET were details of the type(s) of therapy used, therapeutic frequency, type of treatment ending (e.g., planned or unplanned), perceived benefits gained from therapy and areas of improvement, and patient motivation.

Recovery and Improvement. Outcomes were measured using Jacobson and Truax (1991) procedures for establishing reliable and clinically significant change. The Reliable Change Index (RCI) assesses whether pre-post differences are greater than the instruments' measurement error (Jacobson & Truax, 1991). The CORE-OM has an established RCI value of 5 (Barkham, Mellor-Clark, Connell, & Cahill, 2006). Thus, a reduction of 5 points or more indicated reliable improvement, a gain of 5 points or more indicated reliable deterioration, and no change or change less than 5 points was considered no reliable change. Clinically significant change refers to the extent to which post-treatment change is clinically meaningful (Evans, Margison & Barkham, 1998), with clinical improvement defined as movement from the clinical range at intake to the non-clinical (general population) range post-therapy. The recommended clinical cut-off score of 10 was adopted, which has associated sensitivity and specificity values of 87% and 88% respectively (Connell et al., 2007). These criteria allowed patients to be categorised according to four types of change:

v.) *Recovered* (RCSI): comprising patients whose pre-post scores met criteria for both reliable improvement and clinically significant improvement.

vi.) *Improved*: consisting patients whose pre-post score difference met the criteria for reliable improvement only. Included within the category were those patients

who began therapy below the clinical cut-off, who were unable to achieve clinically significant improvement, but could demonstrate reliable change.

vii.) *No Reliable Change*: referring to those patients whose pre-post CORE-OM scores showed no movement or movement that did not reach the reliable change threshold.

viii.) *Deteriorated*: comprising patients who had post therapy scores reliably worse than their intake scores.

Data Analysis

Profiling. Profiling analyses were conducted using SPSS version 21 (IBM Corp., 2012). Due to the large sample size, and therefore high statistical power, statistical significance was determined using an alpha level of .001. Additionally, effect sizes were reported by calculating the difference between means divided by the pooled standard deviation. Effect sizes were described in accordance with conventions outlined by Cohen (1988); .2 (small), .5 (medium), and .8 (large).

Descriptive analyses were used to obtain *profile* characteristics of Voluntary and NHS samples according to: i.) *patient demographics* (age, gender, ethnicity, employment and relationship status, intake severity, and complexity), ii.) *service delivery factors* (types of therapy delivered, dosage, and number of therapies received), and iii.) *therapist caseloads* (caseload size, and aggregated patient intake severity and patient intake risk scores). Descriptive *profiles* between samples were then compared using non-parametric and parametric testing. Proportional differences were examined using chi-squared analyses. Differences in data at the interval/ratio level were tested using t-tests.

Descriptive analyses of patient average outcomes were determined for voluntary and NHS samples. In order to establish an overall treatment effect for each sector, difference in pre- and post-therapy scores were tested using repeated measures t-tests. Patient outcomes were then compared between samples using Analysis of Covariance (ANCOVA), with patient intake score as the covariate.

Multilevel Modelling Analyses. MLM analyses were conducted using MLwiN software v2.29 (Rasbash, Charlton, Browne, Healy, & Cameron, 2009). In total, three models were developed using Iterative Generalised Least Squares (IGLS) estimation procedures: i.) an overall sample model, ii.) a voluntary sector model, and iii.) a NHS sector model. Each model had two levels; patients at level one and therapists at level two, with therapists treated as random effects. Treatment of therapists as random effects assumes therapists are sampled from a larger population and allows for generalisations to be made (Kim et al., 2006). The dependent variable for all models was patient post-therapy CORE-OM scores. Patient and therapist level residuals were examined against quantile-quantile plots in order to assess normality assumptions.

Models were developed in a stepwise fashion according to three stages: i.) single level regression model of patient level variables, ii.) two level random intercepts model, in which therapists were added at level two and their average patient outcome scores allowed to vary about the group average, and iii.) random intercepts and slopes model; as above with the introduction of patient level explanatory variables being allowed to vary between therapists. Thus, a significant random intercept indicates that patient outcomes differ between therapists (i.e., a therapist effect) and a significant random slope suggests the relationship between the patient explanatory variable and outcome variable varies between therapists (i.e., variability between therapist outcomes is due to variation in a given patient factor).

Treatment of Explanatory Variables. For the purposes of MLM analyses, predictive utility of variables was informed by prior research and development of the single level regression model. Variables were entered in the following order: i.) patient variables (initial CORE-OM non-risk score, initial risk score, complexity, age, gender, employment status, relationship status, and ethnicity), ii.) patient level process variables (treatment duration and number of interventions), and iii.) therapist level variables (caseload size, therapist non-risk caseload, and therapist risk caseload). Therapist variables were formed through aggregation of patient level variables. The first model for the full sample contained an additional variable concerning *sector* (i.e., voluntary versus NHS), entered as a patient level process variable, in order to determine the relative effect of the type of mental health provider on outcome.

Continuous variables were entered grand mean centred (Hoffman & Gavin, 1998). A number of categorical variables of interest contained multiple categories, some of which contained relatively few patients. These data can be problematic for MLM procedures, producing less accurate estimates and making model interpretation challenging (e.g., Babyak, 2004). In order to improve the robustness of models, the categories of variables concerning patient ethnicity, employment status, and relationship status were reduced. Data reduction was informed by similarity of category coefficients produced at the single level model stage. Variables were reduced as follows: ethnicity comprised *White British/European* ('White British', 'European', and 'Irish') and *Ethnic Minority* ('Asian', 'Black', 'Chinese', and 'Other'); employment status included *Employed* ('full-time employed', 'part-time employed', 'full-time student', 'retired', and 'other') and *Unemployed/No Formal Work* ('unemployed', 'sick/benefits', 'part-time student', and 'house person'); relationship status consisted *Living with Other* ('single parent', 'partner', 'partner and children', 'carer', and 'other') and *Alone/Non-*

independent Living (‘alone’, ‘temporary accommodation’, ‘shared accommodation’, ‘relatives/friends’, ‘parents/guardians’).

Two categorical variables, presenting problems and type of therapy, could not be included in the models. These variables had a high number of categories which, as previously noted, can be problematic for model robustness. Additionally, these variables could not be meaningfully reduced due to multiple responses being permitted. Therefore, inclusion of factors concerning presenting problems and therapeutic approach were restricted to the number of presenting problems (i.e., complexity) and number of therapies received.

Determining Variable and Model Significance. Significance of explanatory variables was determined by calculating their z-ratio through dividing the resultant variable coefficient by their standard error. Values greater than 1.96 indicated variable significance. Significant variables were assessed for interactions, random intercepts, and random slopes. Variables were additionally tested for polynomial, or curvelinear, relationships with the outcome variable. At each stage of development, differences in the -2*loglikelihood ratio were compared against chi-squared critical values in order to determine model improvement.

Therapist effects for each model were calculated using the Variable Partition Coefficient (VPC), in which the Level 2 unexplained variance is divided by the sum of Level 1 and Level 2 unexplained variance (see formula below, where O^2_{u0} = Level 2 [therapist] unexplained variance and O^2_e = Level 1 [patient] unexplained variance).

$$VPC = \frac{O^2_{u0}}{O^2_{u0} + O^2_e}$$

The derived VPC value was multiplied by 100 to gain the therapist effect.

Ethical Consideration

Ethical approval for the current study was granted by the National Research Ethics Service Yorkshire and Humber (East Leeds) Committee (ref 05/Q1206/128, Amendment 3; Appendix D).

Results

Descriptive profiles per sector are presented first. In order that data are not repeated, descriptive data are provided fully in Tables 2 to 7 alongside comparative statistics within the subsequent comparison section.

Voluntary Sector: Patient Profile

Demographic Information. Voluntary sector patient demographic information is presented in Table 2. Of the total 2,157 voluntary sector patients, 56.1% were female, 89.9% were White British/European, 67.1% were in full- or part-time employment, and 28.2% lived alone. Patients had a mean (SD) age of 38.5 years (10.6; 95% CI = 38.0 - 38.9).

Presenting Problems. Table 3 displays proportions of patients experiencing each presenting problem. Frequently endorsed presenting problems were anxiety (78.8%), interpersonal (70.4%), depression (63.1%), and self-esteem (60.9%). Least common presenting problems were psychosis (.5%), cognitive difficulties (1.3%), eating disorders (1.7%), and personality problems (2.0%). Patient complexity ranged between 1 and 10 presenting problems, with an average (*mdn*) presentation of 4 problems.

Patient Intake Severity. To determine patient intake severity mean clinical CORE-OM, non-risk, and risk scores were calculated (Table 4). Patients had a mean

(SD) CORE-OM score of 17.26 (6.57) average non-risk scores of 20.10 (7.28), and an average risk score of 4.06 (5.54).

Voluntary Sector: Therapeutic Processes

Factors associated with treatment delivery, including therapy frequency and duration, are described in Table 5. Patients attended a mean (SD) number of 13.31 (14.7) sessions delivered most often on a weekly basis (84.1%). Additionally, the proportion of patients receiving each therapeutic approach and the average number of sessions received per therapy were calculated (Table 6). Frequently endorsed approaches were psychodynamic therapy (43%), integrative therapies (37.1%), and person-centred therapy (21.8%). A higher number of sessions were observed across psychoanalytic, systemic, psychodynamic, and integrative approaches. The former two therapies were the least frequently applied interventions.

Voluntary Sector: Therapist Case-Mix

Therapist caseload characteristics were established according to caseload size and patient intake severity. The 50 voluntary sector therapists had caseload sizes ranging from the study minimum of 20 to 167 patients; treating on average 69.64 patients (*SD* = 52.52; 95% CI: 67.4 - 71.9). In terms of patient severity, summarised in Table 7, therapists saw patients with average clinical scores of 17.26 (*SD* = 1.89), non-risk scores of 20.10 (*SD* = 2.06) and risk scores of 4.06 (*SD* = 1.60).

NHS Sector: Patient Profile

Demographic Information. Table 2 summarises demographic information for NHS sector patients. Patients had a mean age of 42.3 years (SD = 13.6; 95% CI: 42.0 - 42.6). Female patients accounted for 73.6% of the sample, ethnic background was

predominantly White British/European (96.1%), 59.4% were employed either full- or part-time, and 50.7% were living with a partner and/or children.

Presenting Problems. The proportion of NHS patients experiencing each presenting problem is shown in Table 3. Anxiety was the most prevalent presenting problem for 81.1% patients followed by depression (72.2%), interpersonal difficulties (54.2%), and self-esteem (50.6%). Least common problems were psychosis (.5%) and cognitive difficulties (1.1%). Patients presented with an average (*mdn*) of 4 problems, ranging between 1 and 11 difficulties.

Patient Intake Severity. NHS patient intake severity was established by calculating mean CORE-OM clinical, non-risk, and risk scores (Table 4). The mean (SD) CORE-OM score was 18.59 (6.26), with average non-risk and risk scores of 21.65 (6.83) and 4.31 (6.07) respectively.

NHS Sector: Therapeutic Processes

NHS patients attended a mean (SD) of 7.52 (8.10) sessions, with less than half (45.5%) receiving therapy on a weekly basis (Table 5). Displayed within Table 6 are the frequencies of therapeutic approaches used and average treatment duration per therapy. Person-centred, integrative, and structured/brief therapies were the most frequently delivered approaches, accounting for 53.8%, 32.3%, and 25.6% of all therapies delivered respectively. Treatments with the longest duration were psychoanalytic and art therapies, which were also the least often provided.

NHS Sector: Therapist Case-Mix

A total of 83 NHS therapists were included. These therapists treated a mean (SD) of 154.51 (82.67; 95% CI: 152.7 – 156.3) patients, with caseload sizes ranging between 20 and 298. Therapist caseload characteristics in terms of average patient

intake severity scores are presented in Table 7. Therapists saw patients with average CORE-OM scores of 18.59 (SD = 1.80), non-risk item scores of 21.65 (SD = 1.84), and risk item scores of 4.31 (SD = 1.97).

Profile Comparisons between Sectors

Demographic Profiles. Comparisons between patient demographic profiles are presented in Table 2. Differences between voluntary and NHS subsample patient demographic characteristics were determined using chi-squared analyses. Analyses revealed significant differences between sectors in terms of gender, ethnicity, employment status, and relationship status ($p<.001$). The voluntary sector was found to have significantly higher proportions of male patients, patients from Black or 'Other' ethnic groups, patients who were unemployed or in full-time employment, and patients who lived alone, with relatives or friends, or in shared/temporary accommodation than the NHS sector. Patients from the NHS sector were more likely to be female, White/European, receiving benefits, retired or a houseperson, and living with a partner and children or with parents/guardians. An independent samples t-test showed the average age of voluntary sector patients ($M = 38.5$, $SD = 10.6$) to be significantly younger than NHS patients ($M = 42.3$, $SD = 13.6$), $t(4248) = -14.11$, $p<.001$), with a small effect size difference of -.29.

Table 2.

Voluntary and NHS Sector Patient Demographic Information and Group Comparisons

Demographic Variable	Voluntary n (%) N=2157	NHS n (%) N=7985	**χ^2 (d.f.)**
Gender:			246.40 (1)*
Male	**947 (43.9)**	2110 (26.4)	
Female	1210 (56.1)	**5875 (73.6)**	
Ethnicity:			205.81 (3)*
White British/European	1939 (89.9)	**7674 (96.1)**	
Asian	48 (2.2)	164 (2.1)	
Black	**61 (2.8)**	54 (.7)	
Other	**109 (5.1)**	90 (1.1)	
Employment Status:			246.16 (8)*
Full-time employment	**1112 (51.6)**	3487 (43.7)	
Part-time employment	334 (15.5)	1252 (15.7)	
Unemployed	**320 (14.8)**	689 (8.6)	
Full-time Student	57 (2.6)	177 (2.2)	
Part-time Student	21 (1.0)	66 (.8)	
Retired	50 (2.3)	**582 (7.3)**	
Receiving Sick/Benefits	127 (5.9)	**774 (9.7)**	
House person	62 (2.9)	**599 (7.5)**	
Other	74 (3.4)	359 (4.2)	
Relationship Status:			258.47 (10)*
Alone	**609 (28.2)**	1592 (19.9)	
Single Parent	244 (11.3)	824 (10.3)	
Partner	502 (23.3)	2020 (25.3)	
Partner and Children	381 (17.7)	**2029 (25.4)**	
Parents/Guardians	99 (4.6)	**651 (8.2)**	
Relatives/Friends	**147 (6.8)**	396 (5.0)	
Carer	17 (.8)	76 (1.0)	
Shared Accommodation	**95 (4.4)**	106 (1.3)	
Temporary Accommodation	**21 (1.0)**	19 (.2)	
Institution/Hospital	-	6 (.1)	
Other	42 (1.9)	**266 (3.3)**	
			t (d.f.)
Mean Age (SD)	38.5 (10.6)	42.3 (13.6)	-14.11*

* $p < .001$

Note: Bold type denotes group with highest proportion following significant pairwise comparison.

Presenting Problems. Differences in sector proportions of patients presenting with each type of problem were observed (Table 3). Chi-squared analyses showed significantly higher proportions of voluntary sector patients experiencing problems of addictions, cognitive difficulties, trauma/abuse, self-esteem, interpersonal, and work/academic ($p<.001$) than NHS patients. However, the voluntary sector had significantly smaller proportions of patients with depression or bereavement/loss difficulties ($p<.001$).

Table 3.

Frequencies of presenting problems in Voluntary and NHS sector patients with group comparisons

Presenting Problem	Voluntary n (%)[a]	NHS n (%)[b]	χ^2 (d.f.)	*p*
Depression	1362 (63.1)	**5765 (72.2)**	66.65 (1)	<.001
Anxiety	1700 (78.8)	6474 (81.1)	5.57 (1)	.018
Psychosis	15 (.7)	39 (.5)	1.37 (1)	.241
Personality Difficulties	95 (4.4)	271 (3.4)	4.98 (1)	.026
Cognitive Difficulties	**50 (2.3)**	91 (1.1)	17.20 (1)	<.001
Physical Difficulties	347 (16.1)	1320 (16.5)	.24 (1)	.622
Eating Disorder	97 (4.5)	318 (4.0)	1.15 (1)	.284
Addictions	**193 (8.9)**	321 (4.0)	85.71 (1)	<.001
Trauma/Abuse	**573 (26.6)**	1698 (21.3)	27.45 (1)	<.001
Bereavement/Loss	538 (24.9)	**3231 (40.5)**	175.20 (1)	<.001
Self-Esteem	**1313 (60.9)**	4038 (50.6)	72.31 (1)	<.001
Interpersonal	**1519 (70.4)**	4327 (54.2)	183.28 (1)	<.001
Living/Welfare	368 (17.1)	1581 (19.8)	8.21 (1)	.004
Work/Academic	**600 (27.8)**	1841 (23.1)	21.06 (1)	<.001

[ab]. Groups were not independent due to patients presenting with multiple problems

Note: Statistical comparisons were based on the presence of an individual disorder; bold type denotes group with highest proportion following significant pairwise comparison.

Patient Intake Severity. The clinical cut-off (≥10) was exceeded by 86.6% of voluntary sector patients and 91.4% of NHS patients, which was a significant between group difference (χ^2 [1, n = 10,142] = 45.72, p<.001). Distributions of CORE-OM total scores per sector are depicted diagrammatically in Figure 2. The box-plots show the somewhat higher average intake CORE-OM score among NHS patients and lower minimum intake scores of voluntary sector patients. However, the overall spread of intake scores, about respective sector averages, is relatively similar.

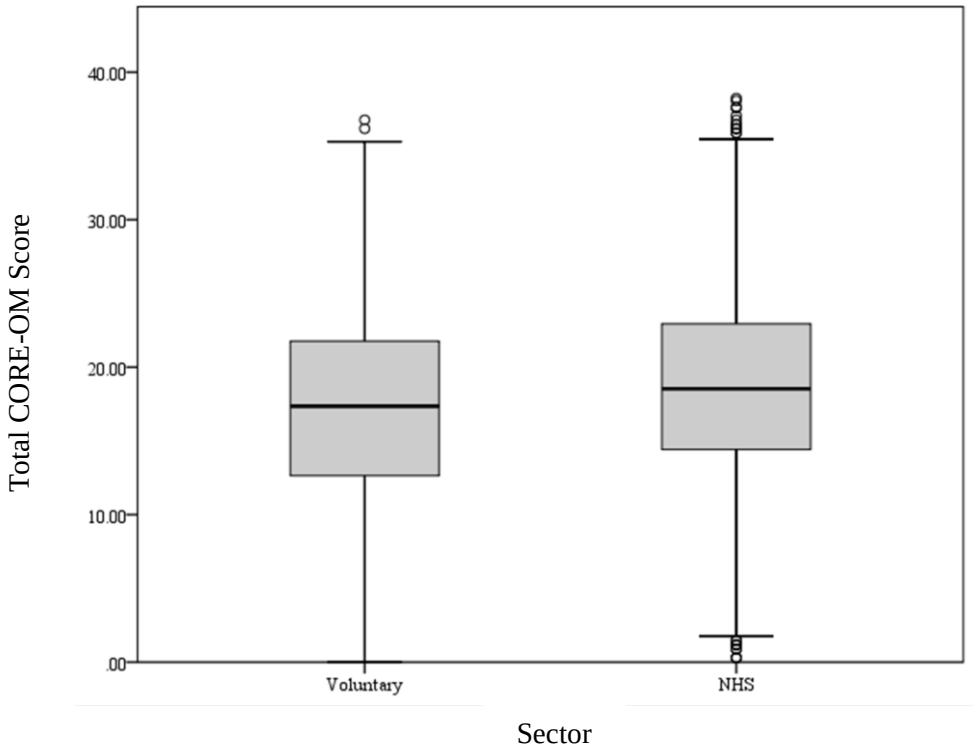

Figure 2. Box-plot distributions of voluntary and NHS sector CORE-OM intake scores

Statistically significant differences were found in CORE-OM clinical and non-risk scores, but not risk scores, between care settings (Table 4). Voluntary sector patients were found to have significantly lower CORE-OM and non-risk scores than NHS patients. However, effect size differences were small (*d* = .21 and *d* = .23 for CORE-OM clinical and non-risk scores respectively).

Table 4.

Intake patient severity according to sector alongside between sector comparisons

Dimension Score	Sector		*t*	df	Effect Size	95% CI
	Voluntary (*M, SD*)	NHS (*M, SD*)				
Total CORE-OM	17.26 (6.57)	**18.59 (6.26)**	-8.42*	3289.94	.21	.16 - .26
Non-Risk	20.10 (7.28)	**21.65 (6.83)**	-8.93*	3254.67	.23	.18 - .28
Risk	4.06 (5.54)	**4.31 (6.06)**	-1.79	3677.14	.03	.01 - .08

* $p < .001$

Note: Bold type denotes group with higher intake severity following significant comparison

Therapeutic Processes. Comparisons of service delivery components revealed significant differences in therapeutic frequency, treatment duration, and the types of therapy provided (Tables 5 and 6). Voluntary sector patients were significantly more likely to be in receipt of treatment on a weekly basis (χ^2 [3, n = 10, 142] = 1057.57, p<.001) and attended significantly more sessions than NHS patients (*M*=13.31, *SD*=14.74 and *M*=7.52, *SD*=8.10 respectively; d=.59, 95% CI: .54 - .63). Significantly higher numbers of voluntary patients received psychodynamic, psychoanalytical, or integrative interventions than NHS patients. Structured/brief, person-centred, supportive, or art therapies were more likely to be provided among NHS settings.

Table 5.

Proportion of patients by therapy frequency and mean number of sessions received

Therapy Frequency	Voluntary Sector n (%)	NHS Sector n (%)	χ^2 (d.f.)
Weekly	1814 (84.1)	3633 (45.5)	1057.73 (3)*
< once weekly	158 (7.3)	3030 (37.9)	
> once weekly	14 (.6)	37 (.5)	
Not at fixed frequency	171 (7.9)	1285 (16.1)	
	Mean (SD)	Mean (SD)	t (d.f.)
Therapy Duration	13.31 (14.74)	7.52 (8.10)	17.55 (2317.35)*

* p <.001

Table 6.

Proportion of patients receiving each therapeutic approach with average treatment durations and between group comparisons

	Voluntary Sector		NHS Sector			
Therapeutic Approach	Frequency[a] n (%)	Mean Duration (SD)	Frequency[b] n (%)	Mean Duration (SD)	χ^2 (d.f.)	*p*
Psychodynamic	**928 (43.0)**	14.92 (17.61)	1216 (15.2)	10.65 (13.15)	786.96 (1)	<.001
Psychoanalytic	**41(1.9)**	33.37 (29.12)	21 (.3)	41.62 (46.03)	74.98 (1)	<.001
Cognitive	105 (4.9)	13.75 (14.89)	536 (6.7)	6.72 (7.34)	9.76 (1)	.002
Behavioural	66 (3.1)	11.80 (11.91)	256 (3.2)	6.20 (2.71)	.12 (1)	.731
Cognitive Behavioural	408 (18.9)	12.98 (12.95)	1313 (16.4)	7.93 (8.22)	7.36 (1)	.007
Structured/Brief	470 (20.2)	7.29 (3.57)	**2041 (25.6)**	6.21 (2.87)	26.77 (1)	<.001
Person-Centred	470 (21.8)	12.16 (11.06)	**4295 (53.8)**	7.09 (6.68)	698.09 (1)	<.001
Integrative	**800 (37.1)**	14.53 (13.42)	2583 (32.3)	6.44 (3.21)	17.17 (1)	<.001
Systemic	38 (1.8)	27.76 (24.90)	188 (2.4)	7.59 (3.48)	2.74 (1)	.098
Supportive	212 (9.8)	14.43 (17.67)	**1457 (18.2)**	6.57 (4.62)	87.54 (1)	<.001
Art	2 (.10)	9.50 (6.36)	**80 (1.0)**	13.33 (28.43)	17.50 (1)	<.001

[ab] Groups were not independent due to patients receiving more than one therapeutic approach

Note: Bold type denotes group with highest proportion following significant pairwise comparison.

Therapist Case-Mix. In terms of therapist caseload characteristics (Table 7), independent samples t-tests indicated significantly smaller caseloads among voluntary sector therapists. Voluntary sector therapist caseloads were characterised by lower CORE-OM, non-risk, and risk scores than NHS therapists (all $p<.001$). However, the effect size analysis revealed therapists to be differentiated by caseload size and caseload non-risk scores, followed by CORE-OM clinical scores. Differences in therapist risk caseloads did not reach the threshold for small effects.

Table 7.

Voluntary and NHS sector therapist case-mix according to patient intake severity

Caseload	Voluntary Mean (SD)	NHS Mean (SD)	t (d.f.)	Effect Size	95% CI
Risk	4.06	4.31	-6.01	.14	.22 - .49
	(1.60)	(1.97)	(4089.35)*		
Non-Risk	20.10	21.65	-31.82	.81	.44 – 1.16
	(2.06)	(1.84)	(3144.91)*		
Clinical	17.26	18.59	-30.10	.71	.35 – 1.07
	(1.89)	(1.80)	(10140)*		
Size	69.64	154.51	-58.09	1.16	.78 – 1.54
	(52.52)	(82.67)	(5359.90)*		

* $p<.001$

Outcomes Analyses

Pre- to Post-Therapy Change. Pre-post therapy scores for each care setting across CORE-OM, non-risk items, and risk items are presented in Table 8 with respective within group effect sizes. For voluntary and NHS care settings, there were significant reductions in overall distress (CORE-OM score), non-risk scores, and risk scores ($p<.001$) post-treatment. Large pre-post treatment effect sizes were achieved by both voluntary and NHS sectors; 1.33 and 1.56 respectively, with smaller but moderate

effect size differences for risk scores. Descriptively, effect sizes across each dimension were higher for NHS services than voluntary services. Analysis of covariance, controlling for pre-therapy scores, revealed significant differences between services in outcomes scores (ANCOVA: $F(1, 10139) = 23.81$, $p < .001$, $r^2 = .19$); with NHS patients having lower average outcome scores.

Reliable and Clinically Significant Change. Patient outcomes were grouped according to four categories: recovered (RCSI), improved (reliable improvement only), no change, and deteriorated. Rates of RCSI were summarised separately for the full and clinical (pre-therapy scores above clinical cut-off) subsamples and are shown in Table 9. For voluntary services, approximately half of patients (50.2%) achieved RCSI, a further 19.7% had reliably improved, 27.7% showed no reliable change, and over 2% reliably deteriorated. The NHS sample demonstrated RCSI rates of 55.8%, with 20.4% showing reliable improvement, 22.7% had no reliable change, and approximately 1% reliably deteriorated. RCSI rates increased to 58.0% and 61.0% for voluntary and NHS settings respectively when considering clinical patients only. Chi-squared analyses ($\chi^2 = 47.97$, $df = 3$, $p < .001$) of the full sample data proportions revealed a higher number of patients achieving recovered status in the NHS sector than voluntary sector. Higher proportions of deterioration and no change were observed in voluntary services, with no differences between groups in patients achieving reliable improvement.

Table 8.

Pre-post treatment outcomes

Care Setting	Dimension	Mean Pre-Therapy (SD)	Mean Post-Therapy (SD)	Mean Difference (SD)	*t*	Effect Size (*d*)
Voluntary (n=2157)	CORE-OM	17.26 (6.57)	8.75 (6.21)	8.51 (6.79)	58.21*	1.33
	Non-Risk	20.10 (7.28)	10.31 (7.04)	9.78 (7.72)	58.85*	1.37
	Risk	4.06 (5.54)	1.48 (3.59)	2.58 (4.91)	24.47*	.55
NHS (n=7985)	CORE-OM	18.59 (6.26)	8.65 (6.46)	9.94 (6.79)	130.86*	1.56
	Non-Risk	21.65 (6.83)	10.23 (7.38)	11.42 (7.73)	132.05*	1.61
	Risk	4.31 (6.07)	1.26 (3.46)	3.05 (5.36)	50.85*	.62

*p <.001

Table 9.

Proportions of Patients meeting four change criteria

Care Setting	Recovered (%)	Improved (%)	No Change (%)	Deterioration (%)
Voluntary				
Full Sample	1083 (50.2)	426 (19.7)	598 (27.7)	50 (2.3)
Clinical Sample (n = 1868)	1083 (58.0)	373 (20.0)	377 (20.2)	35 (1.9)
NHS				
Full Sample	4457 (55.8)	1632 (20.4)	1810 (22.7)	86 (1.1)
Clinical Sample (n = 7301)	4457 (61.0)	1431 (19.6)	1349 (18.5)	64 (.9)

Multilevel Models

Linearity was observed in patient and therapist level residuals following examination of normal distribution curves (Q-Q plots). Therefore, normality was assumed. Each model, in terms of development and associated significant predictors, is discussed first followed by an exploration of key predictors within and between sectors. Therapist effects derived from the models, and therapist variation within sectors, are discussed in separate subsequent sections.

Model 1: Full Sample Model. The primary aim of the full sample model was to determine the relative contribution of the type of mental health provider on patient outcome and therapist effects after controlling for patient and therapist factors. Type of provider was added as a patient process factor. Full model specification is depicted in Figure 5 (Appendix E). The model contained ten main effects, four interaction terms, and five cross-level interactions. Of the main effects, eight concerned patient factors and two related to therapist caseload variables.

Patient demographics factors of *age*, *gender*, *ethnicity*, *employment status*, and *relationship status* were significant predictors of outcome. Patients who were unemployed or living alone/non-independently were found to have worse outcomes, as were patients of ethnic minority groups, male patients, or older patients.

Positive main effects were found for initial patient *non-risk* and *risk* scores, and *complexity*. Higher levels of intake psychological distress, risk, and increased complexity in terms of number of presenting problems contributed to poorer outcomes. Significant positive random slopes were found for both intake *non-risk* and *risk*. The variance function concerning patient *non-risk* suggested there was greater variation between therapists with increased patient *non-risk* severity. The observed positive covariance between therapist intercepts and slopes indicated a fanning out pattern in

which higher patient *non-risk* was of greater detriment to therapists with poorer average outcomes. However, while higher *risk* scores had a similar impact on therapist outcomes, the magnitude of effect was smaller.

Positive interaction terms were observed between *intake non-risk* and *risk* scores, *non-risk* and *age*, and *non-risk* and *employment status*. Combined effects of higher *non-risk* and *risk* scores resulted in poorer patient outcomes, as did higher non-*risk* scores and being older or unemployed/sick. A significant interaction between patient *intake risk severity* and *employment status* indicated that greater risk scores alongside being unemployed/sick was of additional detriment to outcomes.

The addition of process variables suggested an initial significance of *duration*, *number of therapies* received, and *sector*. These initial indicators suggested longer therapeutic duration predicted worse outcomes but higher numbers of therapies were associated with post-therapy improvements. Moreover, a negative main effect of *sector* was indicative of patients attending NHS services having better outcomes than voluntary sector patients. However, all process variables, including sector, became non-significant following control of *therapist non-risk* and *risk caseloads*.

In terms of the latter variables, both *therapist non-risk caseload* and *therapist risk caseload* were predictive of patient outcomes, with differential impacts. The negative main effect of therapist non-risk caseload demonstrated therapists with above average non-risk caseloads had comparatively better outcomes than therapists with lower non-risk caseloads. In contrast, therapists with higher risk caseloads had poorer average outcomes. Therapist caseload size was non-significant.

A number of cross-level interaction terms were observed between therapist caseload factors and patient level variables. *Therapist non-risk caseload* showed negative interaction effects with *patient non-risk severity*, patient *complexity*, and

employment status. These effects indicated that the impact of higher patient non-risk severity, greater complexity, and being unemployed/sick were reduced when such patients work with therapists whose non-risk caseload was higher than average. However, the reverse was true for patients living alone/non-independently. Combined effects of higher average therapist risk caseload and higher patient non-risk severity were also of additional detriment to outcomes.

Model 2: Voluntary Sector Model. Development of the model was to examine the factors, both patient and therapist, contributing to outcomes and to determine the degree of therapist effects within voluntary sector services. Model specification is presented in Figure 6 (Appendix F). The final model contained eight main effects, with two random slopes, and two linear interaction terms. Of the main effects, seven concerned patient factors and the remaining effect related to a process variable.

Significant main effects were observed across patient level predictors with the exception of *complexity*, which did not remain significant once patient employment and relationship status' were accounted for. Poorer outcomes were associated with living alone/non-independently, unemployment/sick, patients who were male or older, or being from an ethnic minority.

The impact of patient intake severity according to *non-risk* and *risk* scores resulted in positive main effects. Thus, more severely distressed patients or patients with elevated levels of risk had poorer outcomes. Additionally, significant random slopes were found for both patient *non-risk* and *risk* intake scores. Examination of respective covariance matrices and variance functions suggested increases in patient intake non-risk or risk scores resulted in more variability in therapist outcomes. Specifically, therapists with poorer outcomes were more adversely affected by higher patient intake severities concerning non-risk and risk.

Significant positive interaction terms were observed between *intake non-risk severity* and *intake risk severity*, and *intake non-risk* and *employment status*. These interaction terms indicated having higher scores on both non-risk and risk items had an additional detrimental impact on outcome, as did greater intake non-risk severity alongside being unemployed/sick.

The addition of therapeutic process variables produced a negative main effect concerning *number of therapies* received; indicating improved outcomes were associated with having multiple therapies. Therapeutic *duration* was non-significant. Therapist caseload factors contributed little to the model and were found to be non-significant. However, *therapist non-risk caseload* and *therapist risk caseload* variables were approaching significance. Therapist risk caseload only became significant when therapist non-risk caseload was included in the model. These findings might be indicative of an existing correlation between therapist non-risk caseload and therapist risk caseload, and suggested an indirect influence of the latter variable on outcomes.

Model 3a: NHS Sample. The model examined the relative contributions of patient and therapist factors toward outcomes and determined the therapist effect amongst NHS services. Figure 7 (Appendix G) presents full model specification. The model included nine main effects, comprising eight patient factors and one process factor, and four interaction terms.

Patient demographics of *age*, *gender*, *ethnicity*, *employment status*, and *relationship status* were significant predictors of outcome. Patients who were unemployed or living alone/non-independently were found to have worse outcomes, as were patients of ethnic minority groups, male patients, or older patients.

Positive main effects were found regarding patient *intake non-risk* and *risk* scores; indicating that increased levels of intake distress or risk resulted in poorer

outcomes. Significant positive random slopes were found concerning patient intake non-risk and risk scores. Inspection of covariance matrices and variance components indicated increases in patient intake non-risk or risk scores resulted in greater variability in therapist outcomes. Poorer therapist outcomes were associated with increased levels of patient intake non-risk severity or risk severity, with greater detrimental effects on those therapists with above average outcomes.

Two positive interaction terms were found concerning patient non-risk scores; comprising patient *non-risk* and risk *scores*, and *non-risk scores* and *employment status*. These interactions suggested having higher scores on both non-risk and risk items had an additional detrimental impact on outcome, as did greater intake non-risk severity alongside being unemployed/sick. Patients with a greater degree of *complexity* (i.e., multiple problems) also had poorer treatment outcomes. The positive linear interaction between *complexity* and initial *non-risk severity* meant that a higher number of presenting problems with a higher intake severity resulted in additional detriment to outcome.

A positive main effect was observed concerning therapeutic *duration*. The effected indicated that longer intervention lengths (i.e., greater number of sessions) were associated with poorer patient outcomes. The negative linear interaction found between patient *intake risk severity* and therapeutic *duration* indicated that where a patient has above average intake risk scores and receives a higher number of sessions, there is a small adjustment towards better outcome. Therapist caseload factors were not found to contribute toward patient outcomes.

Model 3b: NHS Primary Care Sample. Due to the heterogeneity of the NHS sample, and to overcome potential inflation of therapist effects due to unmodelled service level effects, a post-hoc model was produced for the primary care subsample.

The focus on primary care primarily concerned the availability of data, with relatively fewer data contributions derived from secondary and tertiary care services. Full model specification is presented in Figure 8 (Appendix H), which comprised ten main effects, seven linear interaction terms, and one cross-level interaction effect.

Of patient demographic variables, *employment status*, *ethnicity*, *relationship status* and *age* predictors of outcomes, with worse outcomes associated with unemployment/sick, being of a minority group, living alone/non-independently, or being older. Outcomes were not associated with patient gender.

Significant independent associations were found for intake severity in terms of *non-risk*, *risk*, and *complexity*; with non-risk scores demonstrating the strongest association with outcome. Increased levels of intake distress, risk, or complexity resulted in poorer outcomes. Two positive random slopes were observed, one concerning patient intake *non-risk* scores and one of patient *risk* scores, and were indicative of greater variation in therapist average outcomes as a function of increased patient intake distress and risk. Examination of coefficient matrices, variance functions, and prediction plots revealed: i.) higher than average patient non-risk scores were of greater detriment to therapists whose average outcomes were poorer, and ii.) a trend toward therapists with poorer outcomes to be more adversely impacted by increased patient risk, though the magnitude of effect appeared relatively small.

A number of interaction terms were also observed between intake severity factors and patient demographic variables. Patient non-risk scores demonstrated a higher number of interactions comprising *non-risk x risk*, *non-risk x complexity*, *non-risk x age*, and *non-risk x employment status*. The positive interaction terms indicated greater detriment to outcomes where patients have both higher than average non-risk

scores and: i.) higher levels of risk, ii.) present with a greater degree of complexity, iii.) are older, and iv.) are unemployed/sick.

Therapeutic *duration* was a significant predictor, with longer treatment durations associated with poorer outcomes. Combined effects of greater treatment length and higher intake risk severity suggested a small adjustment toward improved outcome. However, both unemployment/sick and receipt of a higher number of sessions was detrimental to outcome (*unemployed/sick x duration* interaction). The number of therapies received was not statistically significant.

Of the therapist caseload factors, *therapist non-risk caseload* and *therapist risk caseload* were independently associated with outcomes, with differential effects. The negative main effect of therapist non-risk caseload demonstrated therapists with above average non-risk caseloads had comparatively better outcomes than therapists with lower non-risk caseloads. In contrast, therapists with higher risk caseloads had poorer average outcomes. Additionally, a negative cross-level interaction between patient intake non-risk severity and therapist non-risk caseload was observed. The cross-level interaction suggested the impact of patient intake non-risk severity on outcome was reduced when such patients work with therapists whose non-risk caseloads are higher than average.

Exploration of Model Predictors

Model development was supplemented with further exploration from which to determine patterns between primary predictor variables and outcomes both within and between sectors. As such, the exploration of patterns was restricted to voluntary, NHS, and primary care models. Table 10 summarises the main and random effects found per model.

Common predictors concerned patient demographic variables and patient intake severity, with the exception of gender and complexity respectively. Table 11 displays mean outcome scores for categorical patient demographic variables according to sector. Employment status demonstrated the greatest difference in outcome scores, with unemployment/sick resulting in outcomes between 2.6 and 3.0 points higher than employed/student status. For voluntary and NHS samples outcome differences were next greatest for ethnicity, in which patients of ethnic minorities had treatment scores 2.7 and 2.3 points higher respectively, with primary care showing larger differences in outcomes between relationship status groups. While gender was not a significant predictor of outcome in primary care, there was a consistent pattern of male patients having poorer outcomes than female patients across sectors. The lack of a significant gender effect potentially reflects the relatively lower proportion of males within the primary care subsample.

Table 10.

Model Summaries of Main and Random Effects

	Voluntary Effects		**NHS Effects**		**Primary Care Effects**	
Variable	Main	Random	Main	Random	Main	Random
Constant	7.48 (.28)	1.29 (.41)	8.02 (.25)	3.97 (.69)	7.58 (.24)	2.91 (.57)
Patient Non-Risk	.28 (.03)	.005 (.01)	.30 (.02)	.007 (.002)	.27 (.02)	.005 (.002)
Patient Risk	.09 (.04)	.04 (.01)	.11 (.02)	.014 (.004)	.10 (.02)	.011 (.004)
Complexity	NS	-	.185 (.042)	-	.165 (.043)	-
Age	.024 (.011)	-	.042 (.004)	-	.044 (.005)	-
Gender	.793 (.273)	-	.404 (.136)	-	NS	-
Ethnicity	1.306 (.396)	-	.847 (.313)	-	.933 (.331)	-
Employment	1.165 (.294)	-	1.268 (.143)	-	1.245 (.147)	-
Relationship	.609 (.243)	-	.678 (.127)	-	.708 (.130)	-
Duration	NS	-	.072 (.010)	-	.095 (.016)	-
Therapies	- .287 (.121)	-	NS	-	NS	-
Therapist Caseload (N)	NS	-	NS	-	NS	-
Therapist Non-Risk	NS	-	NS	-	- .492 (.156)	-
Therapist Risk	NS	-	NS	-	.303 (.152)	-

Table 11.

Post-Therapy Outcome Scores for categorical variables

	Voluntary Sector		NHS Sector		Primary Care	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Employment Status						
Employed	1627	8.04 (5.61)	5857	7.85 (5.85)	5480	7.66 (5.68)
Unemployed	530	10.93 (7.38)	2128	10.84 (7.48)	1889	10.30 (7.20)
Ethnicity						
White/European	1939	8.48 (6.02)	7674	8.56 (6.40)	7102	8.27 (6.17)
BME	218	11.19 (7.26)	311	10.89 (7.56)	267	10.13 (7.02)
Gender						
Female	1210	8.60 (6.45)	5875	8.46 (6.36)	5441	8.23 (6.18)
Male	947	8.96 (5.89)	2110	9.17 (6.70)	1928	8.65 (6.29)
Relationship Status						
Living with Other	1186	8.19 (6.01)	5215	8.13 (6.22)	4920	7.29 (6.02)
Alone/Non-Independent	971	9.44 (6.39)	2770	9.61 (6.79)	2449	9.20 (6.50)

Visual analyses were conducted on graphical data concerning intake non-risk and risk severities. In order to aid identification of data characteristics and patterns, data were smoothed using average outcomes of 5-point categories of non-risk and risk scores. The data for intake non-risk and risk scores per sector, alongside average degrees of change, are illustrated in Figure 9. Overall, graphical data indicated voluntary sector patients having post-therapy scores greater than NHS and primary care patients, with lesser degrees of change. Consistent with model observations, patients experiencing higher levels of distress or risk experienced poorer outcomes. These patients also displayed a tendency toward greater degrees of change. However, somewhat differing patterns emerged at more extreme levels of severity. For instance, within the voluntary sector, the amount of pre-post therapy change began to diminish at the highest levels of non-risk severity (≥ 30) and at risk scores of approximately 25 and above. In contrast, NHS and primary care data revealed sharp improvements in post-therapy scores and rates of change in patients experiencing levels of risk of 35 and above. Given the low numbers of patients contributing toward the highest severity categories, caution is warranted in the interpretation of these findings and may explain the lack of polynomial relationships observed.

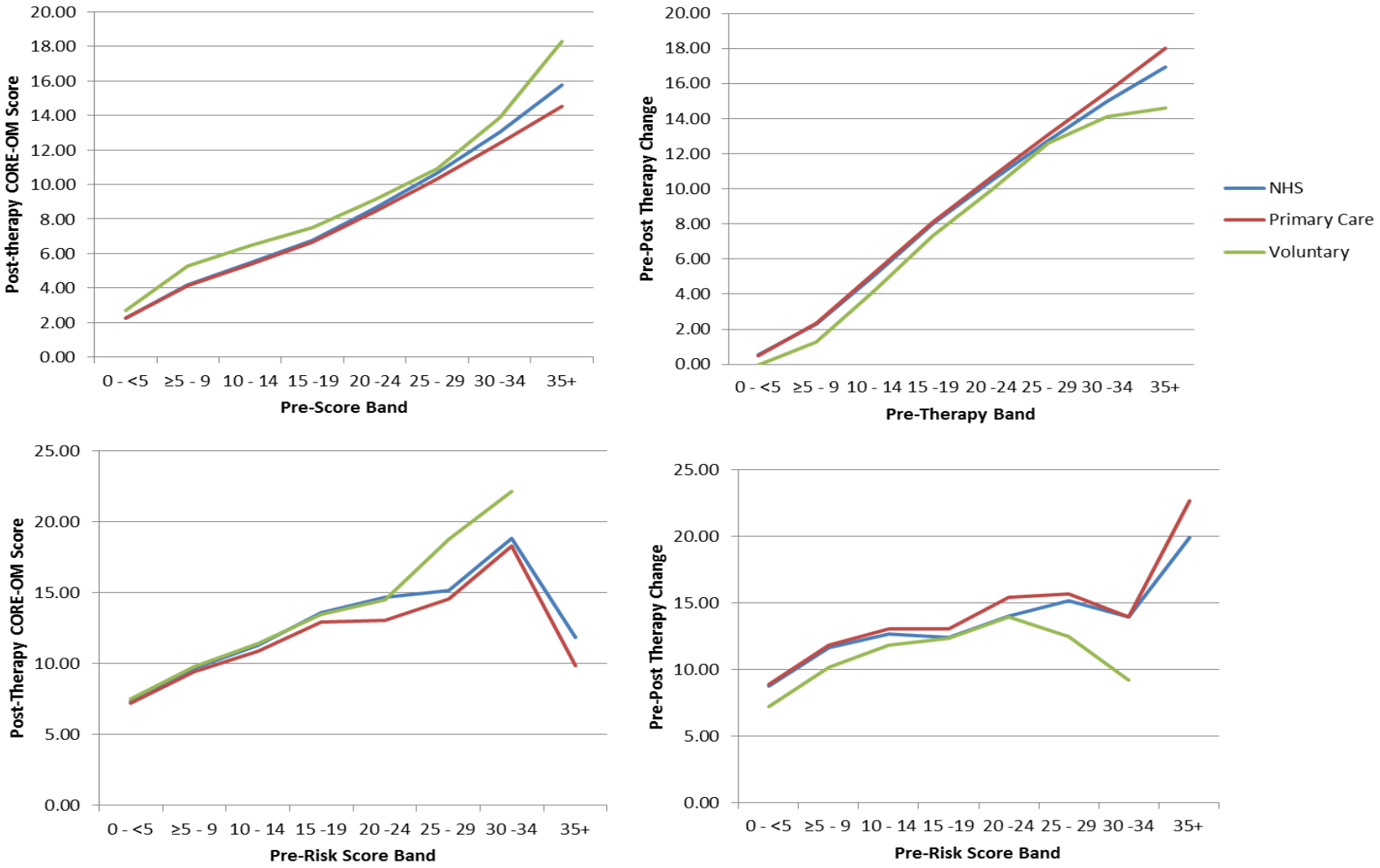

Figure 9. Average post-treatment scores according to pre-therapy score groupings and average degrees of pre-post change

Key differences between models concerned the impact of patient complexity, and process and therapist variables on outcomes. In terms of complexity, a general pattern emerged across sectors in which there was a gradual worsening of outcomes with increasing patient complexity (Figure 10). Despite similarity in trends, voluntary sector outcomes were not significantly impacted by complexity possibly due to worsening outcomes peaking at eight presenting problems prior to improving.

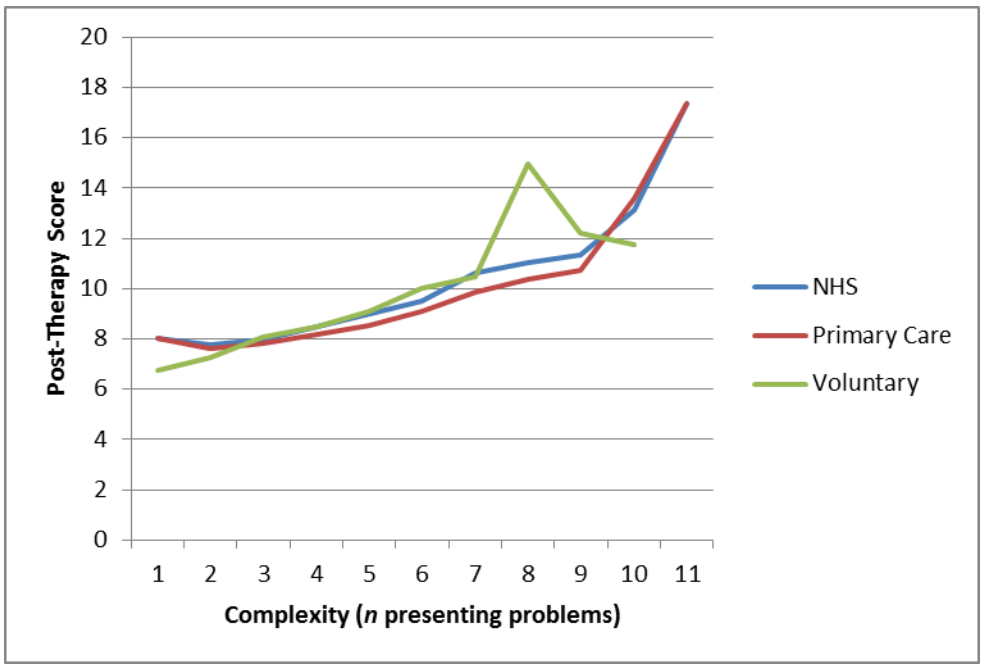

Figure 10. Post outcomes scores according to complexity

Voluntary sector patients were found to benefit from receipt of multiple therapies. The same effect was not observed among NHS and primary care samples. Figure 11 displays average outcome scores according to number of therapies received for voluntary and primary care samples. NHS and primary care outcomes followed the same trajectory and therefore the NHS data were not included for visual clarity. Examination of data indicated sectors differed in the range of number of therapies delivered. The voluntary sector was found to deliver between 1 and 9 therapies

contrasted with 1 to 6 therapies of NHS and primary care settings. Outcomes of the voluntary sector showed gradual improvement with increasing numbers of therapies received. The generally consistent pattern among NHS and primary care settings indicated little systematic change in outcomes according to number of therapies. However, beyond 6 and 4 types of therapy for each sector respectively, there were greater fluctuations in outcomes. The observed outcome variability at maximum numbers of therapies suggests the possibility of an optimum number of therapies for successful outcomes. Alternately stated, too dilute therapeutic approaches may be of less therapeutic benefit.

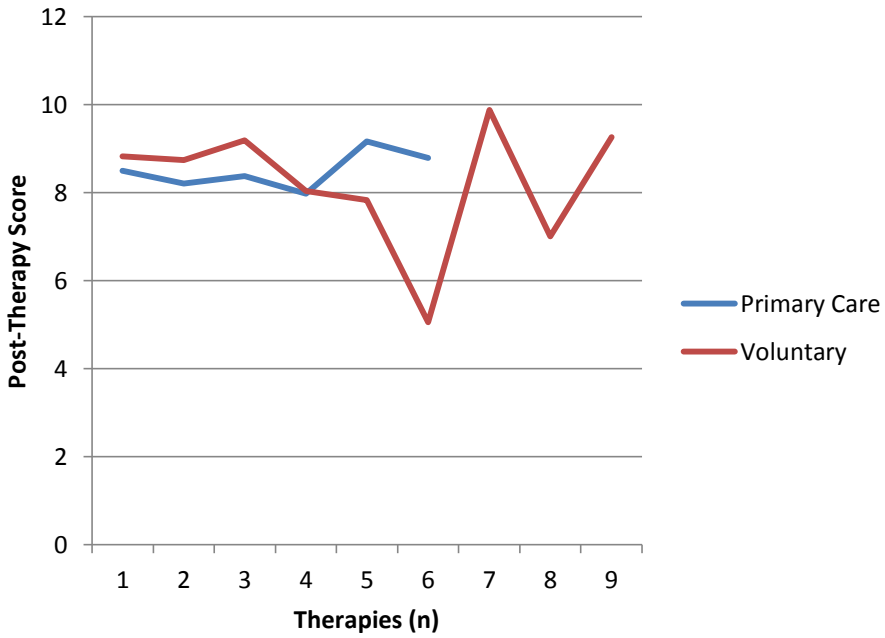

Figure 11. Post-Therapy Outcomes as a function of number of therapies received

Increased therapeutic duration was found to negatively impact outcomes of NHS and primary care patients. Therapeutic duration was not a significant predictor of voluntary sector outcomes. Due to large ranges in attended sessions, analyses were restricted to the outcomes of 2 to 40 sessions, representing the most frequent treatment

lengths, and were categorised according to 3-session increments (Figure 12). For NHS samples, fewer sessions resulted in somewhat better outcomes, with increasing post-therapy outcome scores occurring as a function of sessions attended. Voluntary sector outcomes remained relatively stable across treatment lengths, with variation around the overall outcome mean (8.75, SD = 6.21).

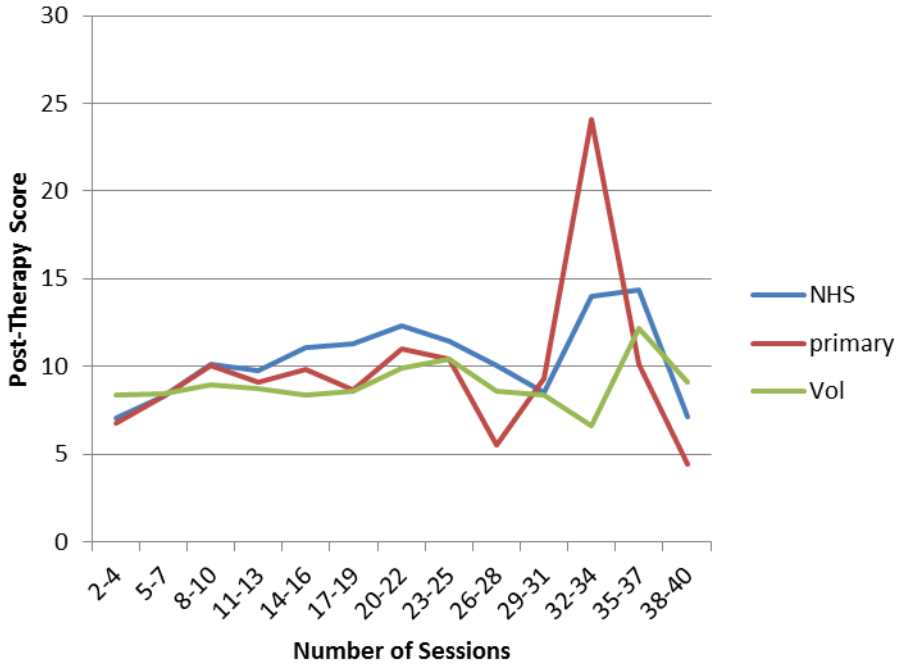

Figure 12. Post-Therapy scores as a function of therapeutic duration

Effects of therapist average non-risk and risk caseloads were significant among primary care settings only. In order to explore such effects, and between sector differences, therapist severity caseloads were categorised into deciles and mean outcomes scores determined (Figure 13). Primary care outcomes demonstrated greater degrees of variation with increasing therapist non-risk and risk caseload severity; potentially accounting for the differences between model specifications. The overall trends suggested poorer outcomes were associated with elevated therapist non-risk caseloads, in contrast to model observations, and higher therapist risk caseloads.

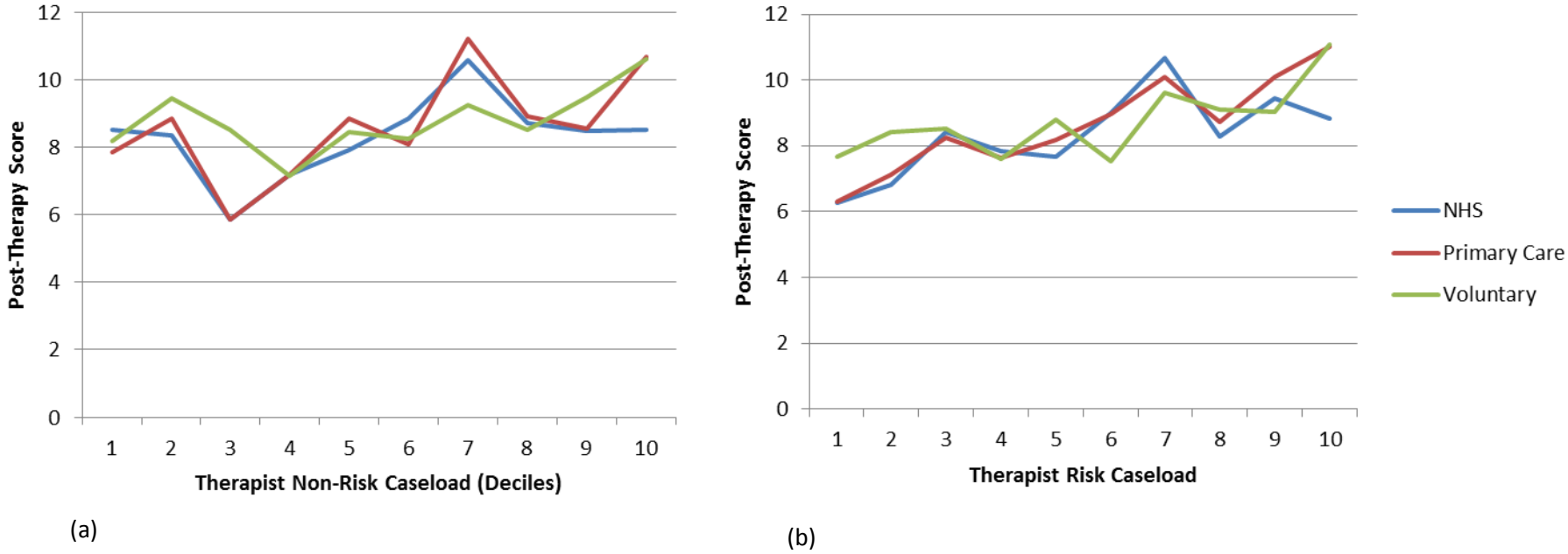

Figure 13. Outcomes according to therapist non-risk and risk caseloads

Therapist Effects

Table 12 displays therapist effects derived from each model. Data are presented according to model development; providing therapist effects for each case-mix model, process model, therapist factors model, and the final model. Significance of the therapist effect is represented by the difference in the -2*loglikelihood and associated z-ratios.

Table 12.

Summary of therapist effects per model according to development

	Model			
Therapist Effect (%)	Full Sample	Voluntary	NHS	Primary Care
Case-Mix	6.52	4.37	12.83	10.53
Process	6.47	4.50	12.61	10.80
Therapist	5.13	4.50	12.61	9.71
Final	4.96	4.49	12.68	9.87
-2*loglikelihood	399*	30*	846*	685*
z-ratio	5.95	3.13	5.77	5.09

*p <.001

Variability of Therapist Effectiveness

Profiles of therapist effectiveness within sectors were considered using caterpillar plots, in which individual therapist residuals were ranked and plotted alongside their respective 95% confidence intervals (Figure 14). Average therapists are represented by confidence intervals crossing zero. Confidence intervals not crossing zero are indicative of above or below average therapists dependent on the sign of the residual. Negative residuals denote above average (i.e., more effective) therapists and

positive residuals represent therapist performance below average (i.e., less effective). The voluntary sector plot identified 45 (90%) therapists of average performance, with 4 (8%) therapists performing below average, and a single therapist (2%) having outcomes above average. Within the primary care sample, 13 (19.7%) therapists were considered more effective with above average outcomes, 36 (54.5%) therapists were of average performance, and 17 (25.8%) therapists were below average. Of the 83 NHS therapists, average performance was found for 36 (43.4%) therapists, 22 (26.5%) were more effective, and 25 (30.1%) were less effective. Descriptively, voluntary sector therapists were more likely to have outcomes within the average range than NHS or primary care therapists. Due to the small numbers of therapists performing above or below average within the voluntary sector, further analyses were not undertaken.

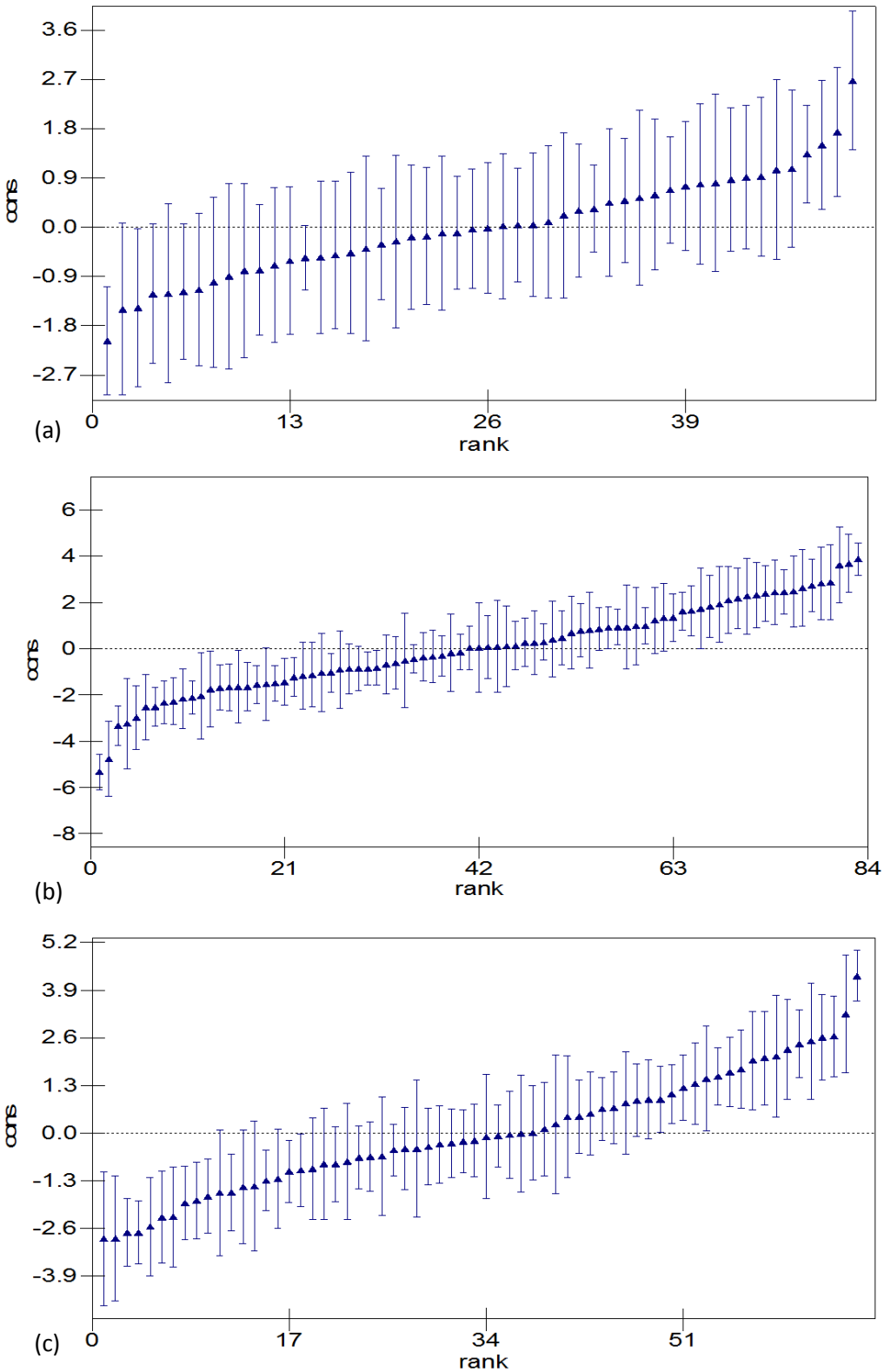

Figure 14. Caterpillar plots of therapist variability: (a) voluntary sector, (b) NHS, (c) primary care. Most effective therapists are presented on the bottom left of each plot

Discussion

The main study findings were as follows: i.) VSO and NHS patients have broadly equivalent clinical presentations and are more likely to differ on sociodemographic variables; ii.) primary differences between sectors concerned their models of service delivery and therapist caseload sizes; iii.) VSO and NHS sectors have comparable outcomes, *sector* was not a significant predictor of outcome; iv.) VSOs had smaller therapist effects (4.5%) than either NHS (12.7%) or primary care (9.9%) samples; v.) VSOs and NHS sectors have shared predictors of outcomes and therapist effects, primarily concerning patient demographic variables and intake severity; and vi.) VSOs and NHS sector outcomes are uniquely affected by variables associated with therapeutic processes and therapist factors.

Service Profile Summary

Profile analyses illustrated key psychosocial characteristics among patients attending VSO and NHS mental health providers alongside respective treatment and service delivery variables. VSO patients, demographically, tended to be female, White British/European, employed, and living alone. Similarly, NHS patients were predominantly female, White British/European and employed, but most frequently living with a partner and/or children. Moreover, VSOs showed comparatively greater proportions of male patients, patients of ethnic minority groups, and patients who were unemployed. Potentially, such differences suggest treatment provider preferences of patients belonging to these demographic groups and reflect the anecdotal evidence concerning VSO unique characteristics in the engagement of underprivileged and underrepresented groups (Appleby, 2009; Miller, 2013).

It was evident across sectors that attending patients experienced a range of presenting problems, with anxiety, depression, and interpersonal difficulties occurring

most frequently. While average intake scores fell within the moderate severity range for both sectors, significant between group differences were found in CORE-OM clinical and non-risk scores. The findings suggested that patients attending VSOs experienced marginally lower levels of distress compared to NHS patients. However, distribution plots showed a marked degree of intake severity overlap between sectors and the small effect size differences observed raise questions of clinical significance. Clinical intake presentations of VSO and NHS patients, therefore, may be broadly equivalent. There were no observed differences in patient levels of risk.

Sectors appeared most discernable in terms of their service delivery processes. VSOs showed a propensity toward providing treatment on a weekly basis, of significantly longer durations, and typically of psychodynamic orientation. In contrast, NHS providers commonly delivered PCT, with shorter treatment lengths, and provided treatment on a weekly basis or less. Additionally, there were differences between sectors in their respective therapist caseloads. While VSO therapists had moderately lower levels of caseload non-risk and overall clinical severity, the most notable difference was caseload size. NHS therapists had significantly larger caseloads. It should be noted that the majority of the NHS sample comprised primary care services and therapists (n=66). Differences in treatment delivery and therapist caseload sizes may therefore be associated with primary care focus on low intensity – high volume models of care (IAPT Mental Health Programme, 2008).

Treatment Outcomes

The effectiveness of VSO and NHS mental health provisions was determined using effect sizes and rates of RCSI. VSO patients were found to have significant improvements in psychological distress following intervention. Large pre-post effect sizes were observed for VSO on CORE-OM clinical scores (1.33) and non-risk items

(1.37), with a moderate effect size for risk items (.55). The CORE-OM clinical score was considerably larger than the finding of Armstrong (2010), who reported a pre-post effect size of .70. Disparity in these outcomes might be attributable to differences in types of intervention and treatment lengths. Armstrong (2010) examined brief solution-focused therapy across an average of four sessions, which appears in contrast of the current study service delivery profile concerning more frequent delivery of psychodynamic approaches and treatment duration of approximately 13 sessions.

Greater magnitudes of change were found for NHS patients across CORE-OM clinical (1.56), non-risk (1.61), and risk scores (.62). Interestingly, the same pattern emerged in which risk items yielded a moderate effect size. Such similarity might reflect the parallels between VSO and NHS patients' intake levels of risk and, given the relatively low levels risk observed, might indicate risk is reasonably contained and therefore not the primary focus of treatment. Between group differences were also found in post-outcome scores after controlling for intake severity, though the difference did not appear clinically meaningful. Within the wider literature, primary care practice-based studies using the CORE-OM have yielded effect sizes between 1.20 and 1.95. VSO outcomes fall within such variability.

In terms of RCSI rates, 58% of VSO patients within the *clinical* subsample met the recovered criterion. While somewhat lower than the rate observed among NHS patients (61%), VSO recovery rates were comparative to outcomes derived in the wider NHS literature (range = 54 – 58%: e.g., Evans, Connell, Barkham, Marshall, Mellor-Clark, 2003; Mullin, Barkham, Mothersole, Bewick, & Kinder, 2006). Improvement rates were equivalent between sectors. Notably, a significantly higher proportion of VSO patients reliably deteriorated. These patients tended to begin treatment within the non-clinical range. One possible interpretation of such a result is that these patients

under-report their initial severity, potentially due to limited insight that is then increased through therapy. However, it is noteworthy that VSOs tended to see more patients falling below clinical cut-off.

The above findings suggest that both VSOs and NHS mental health providers are effective in reducing psychological distress. Moreover, despite VSOs tending to show consistently smaller effect sizes and rates of RCSI, comparison with the wider literature is suggestive that VSO and NHS outcomes are broadly equivalent. Further support for this assertion can be derived from considering the full sample MLM, which indicated that *sector* was not a significant predictor of outcomes following control of patient, process, and therapist factors.

Therapist Effects

The present study is the first to investigate therapist effects within voluntary sector mental health services using a large practice-based dataset and advanced MLM techniques. Therapist effects of 4.5%, 9.9%, and 12.7% were found for VSO, primary care, and NHS samples respectively. The current findings are broadly consistent with previous research findings of therapist effects between 5 and 10% (Crits-Christoph & Mintz, 1991). The higher therapist effect associated with the NHS sample is likely a reflection of un-modelled service effects.

Across all models, variability in therapist effectiveness was influenced by patient intake non-risk and risk severities. Specifically, increased levels of patient non-risk or risk severity were of greater detriment to less effective therapists. These findings suggest that across service providers, patient intake severity has a similar outcome for therapist variability/effectiveness. Previous research also identified differential therapist effectiveness according patient non-risk severity (Saxon & Barkham, 2012). However, the same authors did not find support of a significant therapist effect

regarding patient risk scores. One observation to potentially account for the difference in findings is that patients in the Saxon and Barkham (2012) study averaged somewhat lower risk scores than patients in the present study.

VSOs demonstrated the smallest therapist effects, with the majority of therapists performing within the average range. Prior research has suggested that smaller therapist effects can be associated with manualised therapies (Crits-Christoph & Mintz, 1991). However, it is unlikely that such suggestion provides an adequate explanation given that VSO therapists typically deliver psychodynamic interventions, which are more fluid in approach. A further possibility is level of therapist experience which, in the current context, would suggest VSO therapists have more experience and training. However, information of therapist skills were not available and, therefore, caution is warranted in such interpretation. Patient intake severity is also a potential factor. VSO patients showed a tendency toward lesser intake severity, which would be assumed to result in a smaller therapist effect and the finding is thereby consistent with previous research (e.g. Kim et al., 2006). The magnitude of difference between VSO and NHS patient severity, however, was small. Moreover, patient intake non-risk severity and risk severity were found to increase therapist variability across all models. These findings indicate that relatively small differences in patient severity have the potential to produce much larger therapist effects. Potentially, NHS therapists were more sensitive, or susceptible, to greater degrees of patient severity and, in particular, levels of risk. The latter finding is somewhat evident in consideration of therapist risk caseloads only being significant predictors of outcomes in primary care, who constituted the majority of the NHS sample.

Limitations

A key limitation of the current study concerned data representativeness arising from stages of sample selection. Firstly, the study sample concerned only those patients completing treatment. The extent to which the findings generalise to patients who do not attend or terminate treatment early remains questionable. Completer patients may represent a qualitatively different set of patients than non-completers; potentially inflating overall treatment outcomes and resulting in the underestimation of outcome variability attributable to therapists. However, overcoming issues associated with completer-only samples is challenging when the transferability of intention-to-treat analyses to pre-post PBE designs remains unknown (Barkham et al., 2012). Secondly, high levels of data non-completion introduced possible selective reporting, which was considered through the adoption of a 50% 'good enough' return rate. However, without an appropriately standardised completion rate, issues of 'cherry picking' persist. Lastly, the final sample comprised just 16% of the original, sector specific, dataset; with significant differences emerging between included and excluded patients in respect of demographic characteristics, intake levels of distress, and process factors. Such differences were perhaps due to high statistical power and their clinical significance questionable owing to observed small effect sizes. Nevertheless, caution is warranted in the interpretation and generalisation of findings.

Availability of therapist factors was restricted to those which could be aggregated from existing data. Consequently, the study was unable to examine the specific therapist characteristics that might contribute to patient outcomes or explain the differences in therapist effectiveness both within and between each sector. It is unknown, for instance, whether voluntary sector therapists possess particular attributes that would enable the majority to perform with average degrees of effectiveness or

whether such findings concern their levels and types of training. Additionally, data quality issues concerning type of therapy delivered, alongside a lack of information regarding treatment fidelity, meant that the relative impact of these factors on patient outcomes and therapist effects could not be determined. Therefore, these factors represent potential unknown sources of patient and therapist variation.

Clinical Implications

The present findings provide initial evidence to address policy initiative and quality assessment framework requirements concerning VSO fitness for purpose and suitability as partners in the delivery of mental healthcare (DoH, 2014). VSO effectiveness was broadly equivalent to NHS providers, which would indicate the potential to extend the scope of mental health provisions to VSOs and, in turn, improve access to psychological treatments.

Emerging evidence drawn from profile analyses suggested that expanding the scope of mental health provision to incorporate VSOs might be beneficial to particular groups of patients. VSOs are likely to be attended by more a diverse demographic, particularly in consideration of male and ethnic minority patients. It is possible that these patients find it easier to engage with VSOs or find these services are more responsive to their needs (Appleby, 2009). As such, these types of information have the potential to inform decision making processes in terms of which patients are more likely to engage in and respond to which types of services and their treatment delivery. Similarly, patients who are most likely to benefit from psychodynamic therapy and/or longer treatment lengths might be directed toward VSOs in order to maximise therapeutic gains.

In keeping with previous suggestions, the findings concerning increased therapist variability as a consequence of greater patient intake severities indicate a need

for greater case specific supervision and appropriate allocation of cases to therapists (e.g., Brown et al., 2005; Saxon & Barkham, 2012). Furthermore, given that the largest proportion of VSO therapists performed within the average range of effectiveness might suggest value in peer supervision between VSO and NHS therapists, and would make VSOs true partners in the delivery of mental healthcare.

Research Recommendations

Further research is recommended in order to determine the factors which distinguish differentially effective therapists. Doing so could inform means of supporting and improving the performance of those deemed less effective. Additional understanding is required of the reasons particularly groups of patients attend VSOs and the factors that render the majority of VSO therapists 'average'. Given that the present study is the first to determine therapist effects within VSOs, replication studies will be required. Replication studies would benefit from considering how best to yield a more representative sample.

References

Appleby, L. (2009). *Mental health ten years on: progress on mental health care reform*. London: DoH

Armstrong, J. (2010). How effective are minimally trained/experienced volunteer mental health counsellors? Evaluation of CORE outcome data. *Counselling and Psychotherapy Research*, 10, 22-31. doi: 10.1080/14733140903163284

Arnow, B. A., Steidtmann, D., Blasey, C., Manber, R., Constantino, M. J., Klein, D. N., ... & Kocsis, J. H. (2013). The relationship between the therapeutic alliance and treatment outcome in two distinct psychotherapies for chronic depression. *Journal of consulting and clinical psychology*, *81*, 627-638. doi: 10.1037/a0031530

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). Wiley and Sons.

Barkham, M., Evans, C., Margison, F., & McGrath, G. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health*, 7, 35. doi: 10.1080/09638239818328

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C.… & McGrath, G. (2001). Service profiling and outcomes benchmarking using the COREOM: Towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*, 184-196. doi:10.1037/0022-006X.69.2.184

Barkham, M., Mellor-Clark. J., Connell, J., & Cahill J. (2006). A CORE approach to practice-based evidence: A brief history of the origins and applications of the COREOM and CORE System. *Counselling & Psychotherapy Research, 6,* 3-15. doi:10.1080/14733140600581218

Barkham, M., Mullin, T., Leach, C., Stiles, W. B., & Lucock, M. (2007). Stability of the CORE-OM and BDI-I: Psychometric properties and implications for routine practice. *Psychology & Psychotherapy: Theory, Research & Practice, 80,* 269-278. doi:10.1348/147608306X148048

Brown, G. S., Lambert, J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *The American Journal of Managed Care, 11,* 513-520.

Buckingham, H. (2009). Competition and contracts in the voluntary sector: exploring the implications for homelessness service providers in Southampton. *Policy & Politics, 37,* 235-254.

Care Quality Commission. (2015). *Right here, right now: People's experiences of help, care and support during a mental health crisis*. London: CQC. Retrieved from: http://www.cqc.org.uk/sites/default/files/20150630_righthere_mhcrisiscare_full.pdf

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies.*Journal of consulting and clinical psychology*, *59*, 20-26. doi:10.1037/0022-006X.59.1.20

Curry, N., Mundle, C., Shiel, F., & Weaks, L. (2011). *The voluntary & community sector in health*. London: The King's Fund.

Department of Health. (2006). *Our health, our care, our say: a new direction for community services*. London: DoH. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/272238/6737.pdf

Department of Health. (2007). *New ways of working in mental health*. London: DoH. Retrieved from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_074495.pdf

Department of Health, Mental Health Programme (2008). *Improving Access to Psychological Therapies Implementation Plan: National guidelines for regional delivery*. Department of Health.

Department of Health. (2011). *Extending patient choice of provider*. London: DoH. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216137/dh_128462.pdf

Department of Health. (2012). *No health without mental health: Implementation Framework*. London: DoH. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221687/No-Health-Without-Mental-Health-Implementation-Framework-Report-accessible-version.pdf

Elkin, I., Falconnier, L., Martinovich, Z., & Mahoney, C. (2006). Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Psychotherapy Research, 16,* 144-160. doi:10.1080/10503300500268540

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM. *The British Journal of Psychiatry*, *180*, 51-60. doi: 10.1192/bjp.180.1.51

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1*, 70-72.

Garfield, S. L. (1994). Research on client variables in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.) *Handbook of Psychotherapy and Behaviour Change* (4th ed., pp. 190-228). New York, NY: Wiley.

Green , H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT psychological wellbeing practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy*, 63, 43-54. doi:10.1016/j.brat.2014.08.009

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: Pretreatment patient predictors of outcome. *Clinical Psychology Review*, 22, 875-893. doi:10.1016/S0272-7358(02)00106-X

Hoffmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 23*, 723-744. doi:10.1177/014920639802400504

Hox, J. (1998). Multilevel modeling: When and why. In *Classification, data analysis, and data highways* (pp. 147-154). Springer Berlin Heidelberg.

Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S.W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. Journal of Consulting and *Clinical Psychology, 69*, 747-755. doi:10.1037//0022-006X.69.5.747

IBM Corp . (2012). IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, *59*, 12-19. doi.org/10.1037/0022-006X.59.1.12

Kendall, J., & Knapp, M. (2000). Measuring the performance of voluntary organizations. *Public Management Review*, *2*, 105-132.

Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: a random-effects modelling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research,16*, 161-172. doi:10.1080/10503300500264911

Lambert M., J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. In: Norcross JC, Goldfried MR, (eds). *Handbook of psychotherapy integration*. New York, NY: Basic Books; pp. 94–129.

Lam, A. G., & Sue, S. (2001). Client diversity. *Psychotherapy: Theory, Research, Practice, Training*, *38*, 479-486. doi.org/10.1037/0033-3204.38.4.479

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. (2007). Therapist effects in outpatient psychotherapy: a three level growth curve approach. *Journal of Counseling Psychology*, *54*, 32-39. doi:10.1037/0022-0167.54.1.32

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58,* 127-137. doi: 10.1046/j.0039-0402.2003.00252.x

Macmillan, R. (2011). Supporting the voluntary sector in an age of austerity: the UK coalition government's consultation on improving support for frontline civil society organisations in England. *Voluntary Sector Review*, *2,* 115-124. doi: 10.1332/204080511X560666

Martindale, C. (1978). The therapist-as-fixed-effect fallacy in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 46, 1526-1530. doi:10.1037//0022-006X.46.6.1526

McBride, C., Zuroff, D. C., Ravitz, P., Koestner, R., Moskowitz, D. S., Quilty, L., & Bagby, R. M. (2010). Autonomous and controlled motivation and interpersonal therapy for depression: Moderating role of recurrent depression. *British Journal of Clinical Psychology*, *49,* 529-545. doi:10.1348/014466509X479186

Mellor-Clark, J., & Barkham, M. (2006). The CORE System: Developing and delivering practice-based evidence through quality evaluation. *Handbook of counselling and psychotherapy*, *2*, 207-224.

Miller, R. (2013). Third sector organisations: unique or simply other qualified providers?. *Journal of Public Mental Health*, *12,* 103-113. doi: doi.org/10.1108/JPMH-10-2012-0014

National Health Service for England. (2015). *The five year forward view for mental health*. London: NHS England. Retrieved from: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf

Ostler, K., Thompson, C., Kinmonth, A. L. K., Peveler, R. C., Stevens, L., & Stevens, A. (2001). Influence of socio-economic deprivation on the prevalence and outcome of depression in primary care – The Hampshire Depression Project. *British Journal of Psychiatry, 178,* 12-17. doi:10.1192/bjp.178.1.12

Rasbash, J., Charlton, C., Browne, W. J., Healy, M. and Cameron, B. (2009). MLwiN Version 2.1. [Software]. Available from http://www.bristol.ac.uk/cmm/software/mlwin/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Roth, A., & Fonagy, P. (2006). *What works for whom?: a critical review of psychotherapy research*. New York, NY: Guilford Press.

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80,* 535-546. doi:10.1037/a0028898

Vocisano, C., Klein, D. N., Arnow, B., Rivera, C., Blalock, J. A., Rothbaum, B., . . .Thase, M. E. (2004). Therapist variables that predict symptom change in psychotherapy with chronically depressed outpatients. Psychotherapy, 41, 255-265. doi:10.1037/0033-3204.41.3.255

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914-923. doi:10.1037/0022-006X.73.5.914

Appendix A: Clinical Outcomes in Routine Evaluation-Outcome Measure

<copyright information removed>

Appendix A: Continued

<copyright information removed>

Appendix B. CORE Assessment Form

<copyright information removed>

Appendix B: Continued

<copyright information removed>

Appendix C. CORE End of Therapy Form

<copyright information removed>

Appendix C: Continued

<copyright information removed>

Appendix D. Ethical Approval Letter

NHS
Health Research Authority

NRES Committee Yorkshire & The Humber - Leeds East
Jarrow REC Centre
Room 002
Jarrow Business Centre
Rolling Mill Road
Jarrow
Tyne and Wear
NE32 3DT

Tel: 0191 428 3387

24 February 2014

Professor Michael Barkham
Professor of Clinical and Counselling Psychology
University of Sheffield
Centre for Psychological Services Research
Department of Psychology
Western Bank
Sheffield
S10 2TN

Dear Professor Barkham

Study title:	**An evaluation of the effectiveness of the psychological therapies as delivered in routine practice settings within primary and NHS service settings.**
REC reference:	**05/Q1206/128**
Amendment number:	**Amendment 3, 05/02/14**
Amendment date:	**05 February 2014**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

It was noted that data had been gathered for some time, and clarification was requested as to whether the researchers had, or intended to, disseminate any interim results to participants.

You replied that because the dataset is anonymised, you cannot, de facto, disseminate interim results to participants. You had published extensively on the dataset such that the yield is in the scientific and public domain. Accordingly, you had disseminated findings in the way you envisaged but it would never have been possible to direct these to participants due to anonymity. You added that you could provide a listing of these publications if that would be of assistance to the Committee in arriving at a decision.

In addition to this, it was questioned whether the researchers were still anticipating using their previous plans for dissemination with this current data set, or if they had any new plans as the additional data set would be quite considerable.

You responded that because the dataset was anonymised at the level of the individual participant and the organisation (other than type of organisation), the same restrictions would apply in that you cannot feedback directly to participants at any level. However, the key difference in the additional data set was that the sectors from which the data was drawn, was broader. In that sense, it therefore enabled you to make comparisons between organisational sectors that were not possible in the earlier data set. Hence, while the data set is large in terms

of the total number of participants, when it was clustered according to the differing types of organisations (i.e., NHS [primary, secondary, tertiary], voluntary sector, university counselling, workplace counselling), then the number of participants within each type was considerably reduced. It was this diversity of organisational type that was the rationale for the current data set. In light of this, your dissemination plans would target learned journals and conference presentations focusing on organisational components.

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)	Amendment 3, 05/02/14	05 February 2014

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

05/Q1206/128:	**Please quote this number on all correspondence**

Yours sincerely

pp
Mrs Alison Barraclough
Chair

E-mail: nrescommittee.yorkandhumber-leedseast@nhs.net

A Research Ethics Committee established by the Health Research Authority

Enclosures:	*List of names and professions of members who took part in the review*
Copy to:	*Gill Wells, The University of Sheffield*

NRES Committee Yorkshire & The Humber - Leeds East

Attendance at Sub-Committee of the REC meeting on 20 February 2014

Name	*Profession*	*Capacity*
Mrs Alison Barraclough	Clinical Studies Officer	Lay
Mr Tom Wilson	Consultant ENT Surgeon	Expert

Also in attendance:

Name	*Position (or reason for attending)*
Miss Sarah Grimshaw	REC Manager

Appendix E. Full Sample Model Specification

$$\text{Patient Outcome}_{ij} = \beta_{0j} + \beta_{1j}\text{Patient Non-Risk-gm}_{ij} + \beta_{2j}\text{Patient Risk-gm}_{ij}$$

$$+ 0.005(0.002)\text{Patient Non-Risk-gm}_{ij}.\text{Patient Risk-gm}_{ij}$$

$$+ 0.146(0.034)\text{Complexity-gm}_{ij} + 0.043(0.004)\text{Age-gm}_{ij}$$

$$+ 0.002(0.001)\text{Patient Non-Risk-gm}_{ij}.\text{Age-gm}_{ij} + 0.428(0.125)\text{Male}_{ij}$$

$$+ 1.086(0.249)\text{BME}_{ij} + 1.420(0.132)\text{Unemployed/Sick}_{ij}$$

$$+ 0.087(0.022)\text{Patient Non-Risk-gm}_{ij}.\text{Unemployed/Sick}_{ij}$$

$$+ 0.062(0.023)\ \text{Patient Risk-gm}_{ij}.\text{Unemployed/Sick}_{ij}$$

$$+ 0.804(0.115)\text{Alone/Non-Independent}_{ij} - 0.142(0.059)\text{Therapist Non-Risk-gm}_{ij}$$

$$+ 0.505(0.059)\text{Therapist Risk-gm}_{ij}$$

$$- 0.015(0.006)\text{Therapist Non-Risk-gm}_{ij}.\text{Patient Non-Risk-gm}_{ij}$$

$$- 0.038(0.017)\text{Therapist Non-Risk-gm}_{ij}.\text{Complexity-gm}_{ij}$$

$$- 0.183(0.064)\text{Therapist Non-Risk-gm}_{ij}.\text{Unemployed/Sick}_{ij}$$

$$+ 0.107(0.056)\text{Therapist Non-Risk-gm}_{ij}.\ \text{Alone/Non-Independent}_{ij}$$

$$+ 0.016(0.006)\ \text{Therapist Risk-gm}_{ij}.\text{Patient Non-Risk-gm}_{ij}$$

$$\beta_{0j} = 7.92(0.15) + u_{0j}$$

$$\beta_{1j} = 0.29(0.013) + u_{1j}$$

$$\beta_{2j} = 0.07(0.020) + u_{2j}$$

u_{0j}	$\sim N(0, \Omega_u) : \Omega_u$	1.523(0.256)		
u_{1j}		0.085(0.017)	0.004(0.002)	
u_{2j}		0.026(0.021)	-0.001(0.002)	0.009(0.003)

$$e_{ij} \sim N(0, \sigma^2_e) \quad \sigma^2_e = 29.165(0.415)$$

-2*loglikelihood = 63228.24 (10142 cases)

Figure 5. Full Sample Outcome Model. Coefficient standard errors are shown in parenthesis.

gm = grand mean, *i* = patient ID, *j* = therapist ID

Appendix F. Voluntary Sector Model Specification

Patient Outcome$_{ij}$ = $ß_{0j}$ + $ß_{1j}$Patient Non-Risk-gm$_{ij}$ + $ß_{2j}$Patient Risk-gm$_{ij}$

+ 0.009(0.003)Patient Non-Risk-gm$_{ij}$.Patient Risk-gm$_{ij}$ + 0.024(0.011)Age-gm$_{ij}$

+ 0.793(0.273)Male$_{ij}$ + 1.306(0.396)BME$_{ij}$ + 1.165(0.294)Unemployed/Sick$_{ij}$

+ 0.106(0.039)Patient Non-Risk-gm$_{ij}$.Unemployed/Sick$_{ij}$

+ 0.609(0.243)Alone/Non-Independent$_{ij}$ – 0.287(0.121)TherapiesN-gm$_{ij}$

$ß_{0j}$ = 7.48(0.278) + u_{0j}

$ß_{1j}$ = 0.28(0.025) + u_{1j}

$ß_{2j}$ = 0.09(0.044) + u_{2j}

u_{0j}	~N(0, Ω_u) : Ω_u	1.285(0.410)		
u_{1j}		0.044(0.030)	0.005(0.004)	
u_{2j}		0.105(0.057)	-0.008(0.006)	0.035(0.014)

e_{ij} ~ N(0, σ^2_e) σ^2_e = 27.363(0.856)

-2*loglikelihood = 13338.08 (2157 cases)

Figure 6. Voluntary Sample Outcome Model

gm = grand mean, *i* = Patient ID, *j* =Therapist ID

Appendix G. NHS Model Specification

Patient Outcome$_{ij}$ = $ß_{0j}$ + $ß_{1j}$Patient Non-Risk-gm$_{ij}$ + $ß_{2j}$Patient Risk-gm$_{ij}$

+ 0.004(0.002)Patient Non-Risk-gm$_{ij}$.Patient Risk-gm$_{ij}$

+ 0.185(0.042)Complexity-gm$_{ij}$

+ 0.019(0.005)Patient Non-Risk-gm$_{ij}$.Complexity-gm$_{ij}$ + 0.042(0.004)Age-gm$_{ij}$

+ 0.404(0.136)Male$_{ij}$ + 0.847(0.313)BME$_{ij}$ + 1.268(0.143)Unemployed/Sick$_{ij}$

+ 0.062(0.020)Patient Non-Risk-gm$_{ij}$.Unemployed/Sick$_{ij}$

+ 0.678(0.127)Alone/Non-Independent$_{ij}$ + 0.072(0.010)Duration-gm$_{ij}$

- 0.003(0.001) Patient Risk-gm$_{ij}$.Duration-gm$_{ij}$

$ß_{0j}$ = 8.02(0.245) + u_{0j}

$ß_{1j}$ = 0.30(0.016) + u_{1j}

$ß_{2j}$ = 0.11(0.022) + u_{2j}

u_{0j}	~N(0, Ω_u) : Ω_u	3.969(0.688)		
u_{1j}		0.180(0.036)	0.007(0.002)	
u_{2j}		0.119(0.042)	0.004(0.002)	0.014(0.004)

e_{ij} ~ N(0, σ^2_e) σ^2_e = 27.324(0.438)

-2*loglikelihood = 49320.307 (7985 cases)

Appendix H: Primary Care Model Specification

Patient Outcome$_{ij}$ = $ß_{0j}$ + $ß_{1j}$Patient Non-Risk-gm$_{ij}$ + $ß_{2j}$Patient Risk-gm$_{ij}$

+ 0.005(0.002)Patient Non-Risk-gm$_{ij}$.Patient Risk-gm$_{ij}$

+ 0.165(0.043)Complexity-gm$_{ij}$

+ 0.019(0.005)Patient Non-Risk-gm$_{ij}$.Complexity-gm$_{ij}$

+ 0.044(0.005)Age-gm$_{ij}$

+ 0.002(0.001)Patient Non-Risk-gm$_{ij}$.Age-gm$_{ij}$ + 0.933(0.331)BME$_{ij}$

+ 1.245(0.147)Unemployed/Sick$_{ij}$

+ 0.079(0.021)Patient Non-Risk-gm$_{ij}$.Unemployed/Sick$_{ij}$

+ 0.804(0.115)Alone/Non-Independent$_{ij}$ + .095(.016)Duration-gm$_{ij}$

- .006(.001)Duration-gm$_{ij}$.Patient Risk-gm$_{ij}$

+ .081(.026)Duration-gm$_{ij}$.Unemployed/Sick$_{ij}$

- 0.492(0.156)Therapist Non-Risk-gm$_{ij}$ + 0.303(0.152)Therapist Risk-gm$_{ij}$

– 0.024(0.008)Therapist Non-Risk-gm$_{ij}$.Patient Non-Risk-gm$_{ij}$

$ß_{0j}$ = 7.58(0.236) + u_{0j}

$ß_{1j}$ = 0.28(0.015) + u_{1j}

$ß_{2j}$ = 0.10(0.022) + u_{2j}

u_{0j}	~N(0, Ω_u) : Ω_u	2.909(0.571)		
u_{1j}		0.131(0.030)	0.005(0.002)	
u_{2j}		0.084(0.036)	0.002(0.002)	0.011(0.009)

e_{ij} ~ N(0, σ^2_e) σ^2_e = 26.558(0.442)

-2*loglikelihood = 45260.853 (7369 cases)